Probabilistic Arguments in Mathematics

A Ph.D. Thesis by

Don Berry

University College London

I, Don Berry, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

This thesis addresses a question that emerges naturally from some observations about contemporary mathematical practice. Firstly, mathematicians always demand proof for the acceptance of new results. Secondly, the ability of mathematicians to tell if a discourse gives expression to a proof is less than perfect, and the computers they use are subject to a variety of hardware and software failures. So false results are sometimes accepted, despite insistence on proof. Thirdly, over the past few decades, researchers have also developed a variety of methods that are probabilistic in nature. Even if carried out perfectly, these procedures only yield a conclusion that is very likely to be true.

In some cases, these chances of error are precisely specifiable and can be made as small as desired. The likelihood of an error arising from the inherently uncertain nature of these probabilistic algorithms can therefore be made vanishingly small in comparison to the chances of an error arising when implementing an equivalent deductive algorithm. Moreover, the structure of probabilistic algorithms tends to minimise these Implementation Errors too. So overall, probabilistic methods are sometimes more reliable than deductive ones. This invites the question: 'Are mathematicians rational in continuing to reject these probabilistic methods as a means of establishing mathematical claims?'

Table of Contents

1. Proof and Practice in Mathematics

This chapter is about a rule that characterises contemporary mathematical practice: that mathematicians require proof for the acceptance of mathematical claims. The opening four sections clarify what this entails. In the fifth section I will discuss four desirable features of mathematical practice that this insistence upon proof helps to secure. These are that mathematics has a literature that is both permanent and highly reliable; that there is consensus amongst mathematicians about which results have been definitively established; and that these researchers can in principle find intellectually autonomous reasons for their mathematical beliefs. Lastly, in the final section we see how mathematical practice has changed due to new developments and technologies in recent years.

1.i. Proofs and Proof Presentations

We begin with a rough characterisation of proof, which will suffice to fix a subject matter for discussion. A proof is a finite set of propositions with a particular kind of cumulative inferential structure. It must be possible to arrange the propositions into a sequence such that every proposition is either an axiom or else follows from one or more propositions appearing earlier in the sequence via some accepted mathematical rule of inference (apart from temporary assumptions that are discharged later in the sequence, as in *reductio* arguments). To use the language of graph theory, we can thus arrange the propositions into a finite, rooted, directed tree, where each proposition is represented by a vertex and edges represent implications. This is illustrated by the following diagram of the Cohen Structure Theorem:

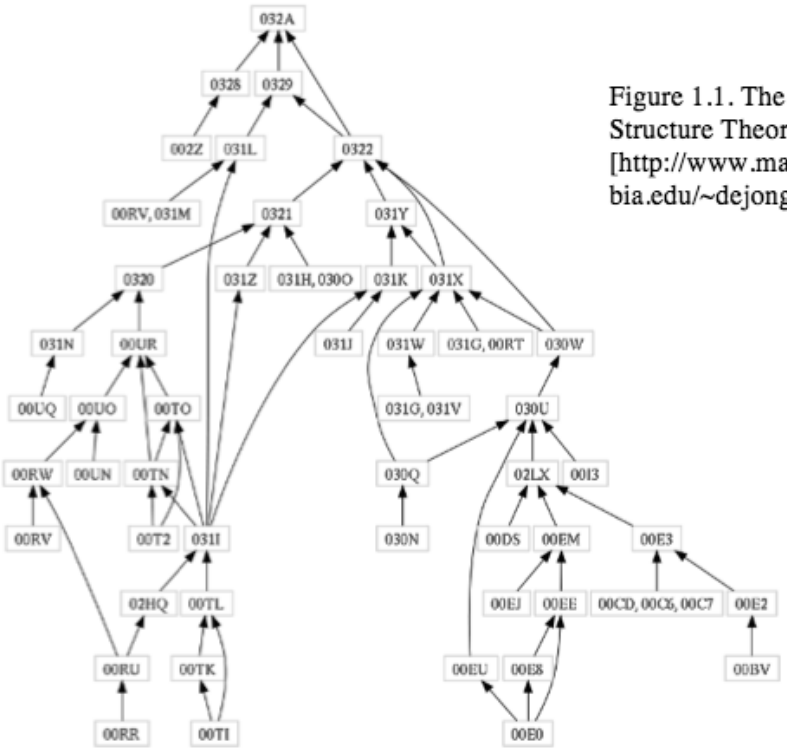

Figure 1.1. The Cohen Structure Theorem. [http://www.math.columbia.edu/~dejong/]

It is clear that if we regard a proof as simply a finite sequence of propositions then some entire sections of argument can be reordered whilst keeping the cumulative inference nature of the argument intact. Yet intuitively we would in many cases still want to regard this as the same proof, despite us now having a different sequence of propositions. Again, this is also true if proofs are identified with directed graphs. However, this individuation need not concern us, and so we will move on to clarifying the other terms in the definition.

The question 'What is a proposition?' is of course a difficult one to answer in itself. We can give at least a partial characterisation: a proposition is the content expressed by a declarative sentence, such as 'There are infinitely many primes'. Propositions are bearers of truth or falsity and are abstract, non-linguistic entities that are the shared objects of propositional attitudes such as belief.[1]

By 'accepted mathematical inference' we shall for now mean the collection of rules that mathematicians acknowledge as acceptable inferences and standardly make use of in the proofs they construct. The question of how this sociological description might be replaced with a more satisfying mathematical formulation will be discussed in Section 5.vi. We shall then see that these rules of inference are truth-preserving, so that it will be impossible for the conclusion to be false if the premisses are true. Knowing the existence of a proof, we are thereby provided with a special kind of *a priori* warrant for believing its conclusion. Once a proof has been found for a result it is known as a 'theorem' thereafter: a significant label that confers upon it the status of having been conclusively established.

In this thesis, proofs thus characterised will be carefully distinguished from what I shall herein call 'proof presentations' (though mathematicians may use the term 'proof' for this concept too). These are the written discourses actually published by mathematicians in journal articles, textbooks and lecture notes. Proof presentations give expression to proofs, and together with lectures are the chief means by which they are communicated. A number of metaphors come to mind for how this is achieved: perceptual, semiotic, cartographic.

A good proof presentation enables any competent reader to know that a proof exists and thus can provide them with strong reasons for believing that its conclusion is a deductive consequence of the axioms or premisses employed. In order to achieve this, proof presentations must make the inferential relations of the propositions they express transparent. It is not enough that they do in fact stand in these relations if this is not clear to a reader. However, the sentences of a proof presentation need not be in one-to-one correspondence with the propositions of the proof it presents. Indeed, proof presentations generally contain gaps, and often make use of subsidiary results that are merely quoted. Where this is the case, we must look at both this proof presentation *and* at presentations of proofs of the cited results to know a full proof of the claim in its entirety.

Proof presentations may comprise a mixture of natural language, mathematical notation and diagrams. The sentences employed can be declarative in character

[1] In this context, we also extend the term 'proposition' to include what is expressed by a sentence containing one or more free variables, such as 'let G be a group'.

('triangles ABC and DEF are similar'), but they may also be in the imperative mood ('construct a tangent to the circle γ meeting the line segment XY at right-angles'). Definitions may be idiosyncratic or delineate entirely new concepts ('we say a group is *Klein-free* if it contains no subgroup isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$'). Proofs of theorems are often achieved by first securing auxiliary results or lemmas which are then drawn together to complete the argument. As mentioned, many of these lemmas are not proved explicitly but quoted from other sources, or are well-known enough that they can simply be stated ('we now invoke the Bolzano-Weierstrass Theorem, showing the existence of a convergent subsequence $x_{j(n)}$'). Corollaries or straightforward consequences of the main theorem are often deduced after it has been proved. Conjectures may be put forward with various degrees of assurance, and informal comments such as suggestions for further applications of derived techniques are often made – though both are sharply distinguished from the central mathematical content of the argument.

In general, then, the claim that mathematicians require proof means that in order for a new result to become accepted (in a sense to be clarified below) a proof presentation must be supplied along with it. In the next section, we get more acquainted with both concepts by looking at some concrete examples of arguments that mathematicians would agree do constitute presentations of proofs.

1.ii. Examples of Proof Presentations

In what follows, the word 'argument' is used to mean either a proof itself or a written discourse that expresses or purports to express a proof. Individual proof presentations are demarcated using boxes.

Theorem 1.2.

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1} \qquad \text{for all } n \geq 2 \text{ and } 1 \leq k \leq n-1$$

The binomial coefficients $\binom{n}{k}$ – read 'n-choose-k' – can be defined as the number of ways of choosing a k-sized subset from an n-set for $n \geq 0$ and $0 \leq k \leq n$. We may also define them using the formula $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, which is easily shown to be extensionally equivalent. Using the latter definition, we can easily prove Theorem 1.2 as follows, using algebra.

$$\binom{n-1}{k} + \binom{n-1}{k-1} = \frac{(n-1)!}{k!\,(n-k-1)!} + \frac{(n-1)!}{(k-1)!\,(n-k)!}$$

$$= \frac{(n-k)(n-1)!}{k!\,(n-k)!} + \frac{k(n-1)!}{k!\,(n-k)!}$$

$$= \frac{n(n-1)!}{k(n-k)!} = \frac{n!}{k(n-k)!} = \binom{n}{k}$$

Although this argument makes use of a number of properties of fractions that are not explicitly mentioned, it is clear that the result in question is unequivocally established, and that it adequately gives expression to a proof. The presentation of a different proof of the same result might run as follows, this time using the set-theoretic definition:

Consider all the k-subsets of a given n set. Let x denote one specific member of the n-set. We organise the k-subsets into two kinds according to whether or not they contain x. The number of k-subsets that do not contain x is equal to $\binom{n-1}{k}$, as we must select all k elements from the remaining $n-1$ elements not identical with x. The number of k-subsets that do contain x is equal to $\binom{n-1}{k-1}$, because this time we need only choose another $(k-1)$ elements to adjoin to $\{x\}$ to make a subset of size k. Hence, the total number of k-subsets of an n-set is $\binom{n-1}{k} + \binom{n-1}{k-1}$, which proves the identity.

The conclusion has again been conclusively established, but this time the character of the proof is much different, and the reliance on natural language in the presentation is far more substantial than for the previous proof. We have also made use of a very different definition for $\binom{n}{k}$, although one that is known to be extensionally equivalent. Thus the same relation between mathematical objects is established, even though reference to them is achieved in a different way.

In the latter proof the reader is not merely following mechanical manipulations of symbols but are invited to use their imaginations to picture a collection of objects that are then manipulated in space and time. These kinds of combinatorial arguments are not foolproof; individual steps may be quite demanding to follow, and one needs to have developed certain skills in order to construct or check them. In this case, the combinatorial argument was also more explanatory: it enabled us to see *why* the two given expressions turn out to be equal.

Sometimes a combinatorial argument may ask us to go even further in conducting a thought-experiment about a practical situation. We illustrate this by looking at a second theorem about binomial coefficients.

Theorem 1.3.

$k\binom{n}{k} = n\binom{n-1}{k-1}$ for all $n \geq 1$ and $1 \leq k \leq n$

Consider a university faculty with n members, $n \geq 1$. We count the number of ways of choosing a committee of k members with one of these members appointed as a chairperson ($1 \leq k \leq n$). We can either choose the whole committee first and then select one of its members to be the chair – which can be done in $k\binom{n}{k}$ ways – or we can first select the chairperson, and then choose the other $(k-1)$ members of the committee from the remaining $(n-1)$ members of the faculty – which can be done in $n\binom{n-1}{k-1}$ ways. These two expressions must therefore be equal, which is precisely the result to be proved.

It will surely be conceded that once again we have found a proof, although it is hard to say exactly how we arrive at this decision. Is the language of the presentation precise enough? Have we paid enough attention to boundary cases, where mistakes often occur because steps are not valid for very small values? For instance, if $k = 0$ the argument does not make sense, because x cannot be a member of the k-subset. More generally, we must always ensure that any values the symbols can take do correspond to a possible instance of the combinatorial interpretation provided.

It is also clear that the details of the particular model I have chosen to base this argument on are irrelevant to the logical structure of the argument itself. The choice of this practical scenario was merely for psychological convenience, and the faculty members individuals that the argument refers to could easily be replaced with mathematical objects if necessary. Nevertheless, I intend all of the arguments given so far to count as proof presentations as they stand. Let us now turn to a third example.

Problem 1.4.

Consider a knockout tournament with 2^n players, where $n \geq 1$. How many matches will be needed to determine the winner?

One obvious approach is to proceed as follows:

There are 2^{n-1} matches in the first round, and half as many in each subsequent round down to the final. Hence, the total number of matches is given by

$$2^{n-1} + 2^{n-2} + \cdots + 2 + 1 = \frac{2^n - 1}{2 - 1} = 2^n - 1.$$

However, consider the following elegant argument:

Each player other than the winner must lose exactly once, and exactly one such loss occurs every match. So the number of matches is $2^n - 1$.

This way of reaching the conclusion might seem to the uninitiated to be less satisfactory than the first because it is purely conceptual, and the argument involves no algebra. Yet it is arguably superior in every way. It doesn't rely on a subsidiary result about the sum of a geometric series, nor an implicit invocation of the principle of mathematical induction. And clearly it can be generalised to other cases far more easily. This shows that a good proof presentation may be very different from how mathematics sometimes appears in the popular imagination, i.e. numerical or symbolic calculation, or the rigid application of formal rules.

Theorem 1.5. (Fermat's Little Theorem[2])

Let p be prime and $a \in \mathbb{Z}^+$. Then $a^p \equiv a \bmod p$

Consider the following argument, taken *verbatim* from Arthur Engel's excellent problem-solving manual.[3]

We have pearls with a colors. From these we make necklaces with exactly p pearls. First, we make a string of pearls. There are a^p different strings. If we throw away the a one-colored strings $a^p - a$ strings will remain. We connect the ends of each string to get necklaces. We find that two strings that differ only by a cyclic permutation of its pearls result in indistinguishable necklaces. But there are p cyclic permutations of p pearls on a string. Hence the number of distinct necklaces is $(a^p - a)/p$. Because of its interpretation this is an integer. So

$$p | a^p - a$$

Some details have been skipped over here. For instance, it is not entirely trivial that a cyclic permutation of a polychromatic string of length p always yields a different string, and is only true when p is prime. The argument also makes less sense for the case $a = 1$, as then there will be no polychromatic strings.

These small omissions are easily fixed; however, there is a deeper reason why the argument as it stands will still be unsatisfactory, and hence not a proof presentation at all. Although the guiding intuition is correct, there is a fundamental

[2] Not to be confused with Fermat's Last Theorem, the proof of which this footnote is sadly too narrow to contain.
[3] Arthur Engel, *Problem-Solving Strategies* (New York: Springer, 1998), 120.

issue with the combinatorial interpretation that structures it. It is true that two strings differing only by a cyclic permutation lead to identical necklaces. But necklaces are also three-dimensional objects that can be picked up and 'turned over'. So strings that are reflections (i.e. strings produced by reversing the order of the pearls) of each other lead to identical necklaces too. Hence the group action on the strings of pearls is dihedral rather than cyclic as asserted.[4] The actual number of distinct necklaces in this case is given by Burnside's Counting Theorem.[5] Again, it clearly requires some skill to avoid such mistakes: Engel himself is a renowned writer of problem-solving manuals.

The argument is again easily fixed by replacing the pearls with less symmetrical objects; alternatively, we can give something like the following argument, which is in essence similar but far more formal and elaborate. It is also more self-contained: it does not make use of the term 'cyclic permutation' and an unproven claim about it, although it does make use of some substantive results such as the division algorithm. This time we argue for a slightly more general version of the theorem than might be yielded by an argument similar to that given above, as now a need not be positive but can take any integer value.

Let p be prime. The theorem is trivial if $a = 0$ or 1. Let a be an integer ≥ 2. Consider the set of finite sequences of length p whose elements come from the set $\{1,2,\dots,a\}$. When constructing a sequence, we must make an a-wise choice for each element, so there are a^p such sequences in total.

There are a sequences consisting of the same repeated number, i.e. $(1,1,1,\dots,1), (2,2,2,\dots,2), \dots, (a,a,a,\dots,a)$. Consider S, the set of such p-tuples that use more than one element from $\{1,2,..,a\}$. Then S has $a^p - a$ elements.

We now proceed to set up an equivalence relation on the members of S that will partition S into equivalence classes, each of size p. There are therefore $(a^p - a)/p$ classes, which proves the theorem for $a \geq 2$, as this number must be an integer.
Define a function f from S onto itself as follows:

$$f : \mathrm{S} \to \mathrm{S}$$

$$f\left((a_1, a_2, a_3, \dots, a_p)\right) = (a_p, a_1, a_2, \dots, a_{p-1}).$$

---

[4] See also: Carl Pomerance and Richard Crandal, *Prime Numbers: A Computational Perspective* (New York: Springer, 2001), 150. The authors give the same argument, mentioning 'visual highlights'. The argument apparently originates with Solomon Golomb, who does not commit this error and discusses the possibility of flipping the necklaces over. See Solomon Golomb, "A Combinatorial Proof of Fermat's Little Theorem", *The American Mathematical Monthly* 63 (1956).

[5] Golomb now gives the different answer of $\frac{a^p - a}{2p}$, apparently ignoring the possibility that a string might be its own inverse. Golomb, "A Combinatorial Proof of Fermat's Little Theorem".

The function takes a sequence and shifts each element along a place. We note that f is a bijection, as it is easy to find the preimage of any member of S explicitly, and also straightforward to show that f is injective. Hence, there is an inverse function f^{-1}.

We define f^n and f^{-n} as the composition of the function n times and its inverse n times, respectively. We also define f^0 as the identity function on S. If s is any member of S, then $f^p(s) = s$, so $f^{bp+d}(s) = f^d(s)$ for any integers b and d.

We claim that for each sequence s in S, the sequences $f^0(s), f^1(s), \dots, f^{p-1}(s)$ are distinct. Suppose not. Then we have distinct integers i, j satisfying $0 \leq i, j \leq p-1$ such that $f^i(s) = f^j(s)$. Without loss of generality, suppose $i < j$. Then $f^{j-i}(s) = f^{i-i}(s) = f^0(s) = s$.

Now let d be the smallest positive integer such that $f^d(s) = s$. Clearly $d < p$, as $(j - i)$ has this property. Using the division algorithm, we can write $p = kd + b$ for some positive integer k and some integer b with $0 \leq b < d$.

Next we show that $b \neq 0$. Suppose not; then $p = kd$, and either k or d would have to be equal to 1, as p is prime. But this cannot be the case: d cannot be 1, as then s would be of the form $(x, x, x, \dots, x)$, and k cannot be 1, as $d < p$. So $b \neq 0$ and hence $0 < b < d$.

However, now we have that $f^b(s) = f^{p-kd}(s) = f^p(f^{-kd}(s)) = s$, contradicting the definition of d as minimal. So $f^0(s), f^1(s), \dots, f^{p-1}(s)$ are indeed distinct, as claimed.

Now let $[s]$ denote the set $\{f^0(s), f^1(s), \dots, f^{p-1}(s)\}$ and define a relation R on S by rRs being true just when r is a member of $[s]$. We prove that R is an equivalence relation and thus partitions S into equivalence classes of size p, as we have seen that there are p members of $[s]$ for any s in S.

Firstly, we always have sRs, as $f^0(s) = s$ by definition, so R is reflexive.

Secondly, rRs implies that sRr. Assume r and s are distinct; then if $r = f^n(s)$ for some $1 \leq n \leq p-1$ then $s = f^{p-n}(r)$ with $1 \leq p-n \leq p-1$. The case where $r = s$ has already been proved above. So R is also symmetric.

Lastly, we prove that R is transitive. Suppose rRs and sRt. Then there are integers m and n such that $0 \leq m, n \leq p-1$ and $r = f^m(s)$ and $s = f^n(t)$. Let $(n + m) = kp + d$ where k is either 0 or 1 and d is a non-negative integer less than p. Then $r = f^d(t)$, so that rRt.

So R is reflexive, symmetric and transitive, and is indeed an equivalence relation; therefore $a^p \equiv a \bmod p$ whenever $a \geq 0$ and p is prime.

Now fix $a < 0$ and set $b = -a > 0$. Let p be a prime other than 2. Then p is odd, and so $a^p \equiv (-b)^p \equiv -b^p \equiv -b \equiv a \bmod p$. Lastly, we have that

$a^2 - a = a(a-1)$, which is necessarily a multiple of 2, as either a or $(a-1)$ must be even. So $a^p \equiv a \bmod p$ for all primes p and integers a.

This argument is undoubtedly satisfactory from an epistemic perspective, although perhaps too much detail has been given and a more economical presentation relying on previously established results (such as the Orbit-Stabiliser Theorem or Lagrange's Theorem) would have been preferable. Lastly, consider one final argument, which falls short of counting as a proof presentation.

Problem 1.6.

Solve the equation $x^3 - x = 0$

$$\begin{aligned} & x^3 - x = 0 \\ \Longrightarrow\ & x^3 = x \\ \Longrightarrow\ & x^2 = 1 \\ \Longrightarrow\ & x = \pm 1 \end{aligned}$$

Obviously, this fails to be an adequate presentation of the claim that the stated solutions are exhaustive because the derivation of the third line assumes that $x \neq 0$. But $x = 0$ is in fact a solution to the equation, and so the given answer is incomplete. Note however that the argument would not be acceptable even if by dint of good luck this kind of carelessness had not prevented us from arriving at the correct conclusion. Even if the propositions making up the proof were all true, they would not be clearly arranged into a suitable inferential structure.

The examples considered in this section illustrate how proofs can be presented in different formats that make use of diverse kinds of reasoning: algebraic, combinatorial, conceptual. We can also distinguish between proof presentations which leave gaps that a reader can reasonably be expected to fill in by themselves, and arguments which fail to be proof presentations because they contain or rely upon a claim which is in fact false. This was true of Engle's argument, even though it can easily be fixed. In some cases, however, repairing the argument may not be possible: it may be that there is no proof available that the proof presentation-like discourse comes close to expressing. We may compare this to hallucinations that seem subjectively like veridical perceptual experiences but have no genuine objects. The last example also illustrates how a proof presentation must also cover all cases relevant to the conditions given in the theorem. This includes borderline and unusual cases, which often require individual attention.

1.iii. Acceptance and Publication Requirements

In the previous section, we saw some examples of discourses that mathematicians would agree count as proof presentations. Where such agreement is present mathematicians are rationally obliged to come to believe the corresponding conclusion, because the inferential structure of a proofs is always truth preserving. In this section, I will argue that, in a sense of 'accept' to be explained presently, mathematicians will *only* come to accept mathematical results if a proof presentation regarded as adequate has first been given. In the thesis, I make use of two distinct understandings of the term 'acceptance'.[6]

Private Acceptance

This term indicates personal belief on behalf of individual mathematicians, whether or not this belief is announced publicly.

Public Acceptance

This term indicates that a result is an established theorem that is now eligible for unqualified assertion in peer-reviewed journals and other serious mathematical publications such as textbooks, monographs, and edited collections.

In making the claim that mathematicians require proofs for the acceptance of results, I only intend this to be taken in the sense of Public Acceptance. Understood in terms of Private Acceptance, it would in fact be false. In the next chapter (Section 2.v) I will give a discussion of the famous Goldbach Conjecture: the claim that every even integer greater than 2 can be expressed as the sum of two (not necessarily distinct) primes. As we shall see, many mathematicians believe this conjecture on the basis of the vast amount of inductive evidence in its favour, even though a proof has not yet been found. Hence, mathematicians do sometimes Privately Accept results in the absence of proof.

Let us now consider the concept of Public Acceptance in more detail. If a claim is Publicly Accepted then it becomes regarded as suitable for unqualified assertion in peer-reviewed journals of mathematics and other serious academic publications – for example, as an auxiliary result for proving another theorem.[7] The fact that a claim is asserted in a specifically unqualified way means that – if the result is a significant one – it is likely to be circulated and reasserted again in print with the same lack of qualification. It will perhaps also find its way into textbooks and eventually be taught to large numbers of students. As confidence in the result increases, it is inducted into the body of established mathematics, and regarded as secure and certain knowledge. The conclusions of subsequent papers whose arguments rely the asserted result are also seen as true *simpliciter*.

---

[6] As with other such phrases that will be introduced later, I will capitalise the labels I give these and their cognates throughout the thesis, to indicate that a special, semi-technical sense is intended.

[7] We focus on the journal article herein, as this is now the primary vehicle for new mathematical research.

The claim being made, then, is that in order to unqualifiedly assert a mathematical proposition in an article published in a mathematical journal such as *Annals of Mathematics*, *Journal of the American Mathematical Society*, and *Acta Mathematica*, an adequate proof presentation must have been given somewhere in the literature (in the same paper if it is a new result). Claims for which this is not true must be clearly demarcated as a conjecture or informal comment or prediction, and thus not part of the central mathematical progression of the article.

This claim about mathematical practice is partly institutional; concerning the standards maintained by the current major academic journals through which new mathematical research is disseminated. Yet it has also long been reflective of the feelings of mathematicians themselves. Euler writes, 'we should take great care not to accept as true such properties of the numbers which we have discovered by observation and which are supported by induction alone',[8] and Frege after him that 'in mathematics a mere moral conviction, supported by a mass of successful applications, is not good enough.'[9] More recently, Michael De Villiers writes that 'Nobody, today, can really be considered mathematically educated or literate, if he or she is not aware of the insufficiency of quasi-empirical evidence to guarantee truth in mathematics, no matter how convincing that evidence may seem.'[10] Indeed, journals referees and research mathematicians may on occasion be the same groups of people.

Before moving on, a minor development in how mathematical research is communicated should be mentioned. Consider the academic journal *Experimental Mathematics*, founded in 1992 by David Epstein, Silvio Levy and Klaus Peters. This journal caused controversy when it was founded because it allows researchers to publish results for which proofs have not yet been found, on the basis of non-deductive arguments alone. However, in the current publication guidelines given on the journal's website, the founders explain that they do value proof. In fact, the early publication of as yet unproved results is partly intended to facilitate the process of finding proofs, because other researchers, who may be better positioned, can join in the search.[11] Moreover, the unproved results published by the journal are not intended to be regarded as theorems. So the founding of this journal and similar developments since then present no problems with the claims made earlier about the necessity of proof for the Public Acceptance of new mathematical results.

1.iv. Acceptability Conditions for Journal Articles

I have claimed that mathematicians require proof for the Public Acceptance of mathematical claims, though we have not yet given clear criteria for when a given body of discourse is to count as a proof presentation. Indeed, we have already seen

---

[8] Quoted in George Pólya, *Mathematics and Plausible Reasoning, Volume I* (Princeton: Princeton University Press, 1954), 3.
[9] Gottlob Frege, *Foundations of Arithmetic*, trans. John Austin (Oxford: Basil Blackwell, 1980) 1.
[10] Michael De Villiers, "The Role and Function of Quasi-Empirical Methods in Mathematics", *Canadian Journal of Science, Mathematics and Technology Education* 4 (2004): 412
[11] "Statement of Philosophy and Publishing Criteria", accessed June 22nd, 2015, http://www.emis.de/journals/EM/expmath/philosophy.html

that the range of devices mathematicians may employ in their proof presentations is highly varied. So it is difficult to give a precise characterisation of what proof presentations are that can be applied operationally to particular cases. However, in order to explicate our claim about mathematical practice further it will not be necessary to do so. Rather than asking the metaphysical question 'When *is* a piece of discourse a proof presentation?', for our philosophical purposes we need only consider the more modest question 'When is a discourse offered as a proof presentation *regarded as acceptable* for publication in a peer-reviewed journal?'. This suffices to characterise fully the aspect of practice central to the primary question that this thesis will address.

The question of when a piece of discourse is acceptable for journal publication is still a complex one, and any adequate answer will have both descriptive and normative dimensions. For an article to be accepted for publication, a reviewer must judge the article to be capable of convincing any competent reader that a proof exists by means of reasoning explicitly given in the article itself, together with references made to other publications ('competence' here includes awareness of results regarded as so widely known within this area of mathematics to not require separate justification or referencing). This is only a minimal constraint, however, and will need to be supplemented with other criteria.

Let us consider things from the perspective of the researcher. As with any academic discipline, the process of learning how to write a journal article generally relies on familiarity with a number of paradigm examples of what a good article looks like, together with experience gleaned from receiving several rounds of feedback on initial attempts. Of course, by the time a young researcher is in the position to submit articles to this kind of publication they will have constructed a great number of proof presentations in other increasingly advanced mathematical contexts throughout the course of their academic careers: geometry homework at secondary school, A level examination questions in algebra, proof questions in example sheets on an undergraduate degree, or qualifying exams after their first year of graduate study. We consider how the relevant standards are determined across these different contexts, with journal articles regarded as a special case. I will discuss four dimensions of variation: level of rigour, level of formality, gaps left unfilled, and subsidiary results relied upon.

The concept of rigour will be discussed more fully in Chapter 5, but we give a brief account here as well. Rigour is largely a matter of three things: making use of a precise (if implicit) understanding of how the mathematical objects, properties and operations involved in the proof presentation are being defined; ensuring that any assertions made about them are categorically true and do stand in the intended place in the logical structure of the argument; and paying attention to all possible cases – even if they seem unimportant, unusual, or of little interest. Intuitively, if we imagine the written argument as a map of (perhaps only part of) the underlying proof, it is rigorous if it represents every piece of the terrain and does so with a high degree of accuracy. Hence, rigour is a question of how adequately a proof is expressed in our written communication.

Generally speaking, the level of rigour required from a student in their written mathematics will increase throughout their education until a high point rigour is

reached, perhaps during a first course in analysis. Here expressions such as $\lim_{n\to\infty} x_n = 2$ first receive clear definition and, for example, results are established for all functions whatever rather than merely those that behave as we would intuitively expect (see Section 5.v). We should distinguish these purely educational contexts, where the individual checking the proof already knows the answer is true and stringent standards of rigour are required largely for the edification of the student, from those front-line research contexts where the result may be entirely new, or a conjecture that is not yet Publicly Accepted.

A less rigorous proof presentation may suffice to convince a reviewer that a new result is true: for example, it may leave out a case that she can deal with herself. But if a reviewer is to approve only arguments that will be found convincing by the entire readership of the journal, and facilitate the Public Acceptance of a result, she must insist upon the core structure of each argument being presented in an incontrovertible and fully rigorous manner. Consider the following passage, taken from an A level textbook:

> 'For an increasing function in the interval (a, b), if x_1 and x_2 are two values of x in the interval $a \leq x \leq b$ and if $x_1 < x_2$ then $f(x_1) < f(x_2)$.
>
> It follows that $f'(x) > 0$ in the interval $a \leq x \leq b$.'[12]

There is a notational confusion between open and closed intervals here, and in any case the property defined – usually known as 'strictly increasing', rather than 'increasing' *simpliciter* – only guarantees the derivative is non-negative throughout the interval. $f(x) = x^3$ is a simple counterexample to the claim as stated. As mathematics is a subject that lends itself to very precise expression, one feels that this sloppiness is a shame even here. But within a research context this would certainly not be acceptable in the statement of a result. In pure analysis, a mathematician would also be careful to explicitly restrict the scope of the theorem to functions that are differentiable throughout the interval before the statement of the main result was given.

Even in published journal articles a lack of rigour may occur in some of the details, however. Papers are often published where notation is abused, such as using *x* as both a dummy variable of integration and a fixed integral limit. Symbols such as 'dx' or '$1/\infty$' are employed despite being given no coherent or unambiguous interpretation. And researchers may miss out cases that are seen as physically impossible or merely unimportant, especially within applied mathematics. These inaccuracies are generally considered undesirable, although the occurrence of one or two such lapses may not threaten the acceptability of an entire article, assuming they are easily corrected and do not undermine the integrity of the central line of argument.

Formality is related to rigour but not to be identified with it, for it is often more a matter of expositional style. In a formal proof presentation, the precision with

[12] Keith Pledger and Dave Wilkins, *Edexcel AS and A Level Modular Mathematics, C2* (Portsmouth: Heinemann, 2008), 142.

which the mathematical objects are identified in rigorous proof presentations is extended more widely across the vocabulary of the whole discourse. This includes the parts that have only an explanatory purpose. Compare the two arguments given for Theorem 1.2; the combinatorial proof presentation is less formal because it is written in natural language, using terms such as 'consider', 'select' and 'choose' which describe intentional human actions that are not inherently mathematical in nature. The first proof presentation uses only algebra and so is more formal. Yet both arguments are fully rigorous; likewise the proof of Theorem 1.3 and the second solution to Problem 1.4, which are given even less formal expositions. Complete formality would be achieved by writing in a fully regimented style according to strict grammatical rules stipulated recursively, similar to a derivation in a formal system, in the sense of mathematical logic (see Section 5.vi).

We should be wary of assuming that in more sophisticated contexts a higher level of formality is always required. Indeed, the example of the two proofs of Theorem 1.2 show that sometimes a less formal proof with more explanatory value is preferable. Generally speaking, a high level of formality is not a necessity prerequisite for journal articles. The following opening paragraph from a journal article – written in informal, interrogative language – is fairly typical:

> 'Fix positive integers m, n and let f be a real-valued function defined on an (arbitrary) given subset $E \subset \mathbb{R}^n$. How can we tell whether f extends to a C^m function F on the whole $\mathbb{R}^n$? If such an F exists, then how small can we take its C^m-norm? What can we say about the derivatives $\partial^\alpha F(x)$ at a given point x? Can we take F to depend linearly on f?
>
> Suppose E is finite. Can we compute an extension F whose C^m-norm has the least possible order of magnitude? How many computer operations does it take?'[13]

Gaps in proof presentations occur either when a proposition that constitutes part of the corresponding proof fails to receive individual expression within the discourse itself, or when explanation is left out for steps in the argument that require it. Let us consider the former kind of gap first. We have said that for a journal article all cases should be taken into account, and clearly the individual lemmas that form the high points of the proof – that is, the conclusions pertaining to discrete segments of its structure, in the sense of the Cohen Structure Theorem pictured in Diagram 1.1 – should certainly be clearly stated. But on the way to establishing these auxiliary claims, not all the steps along the way need receive explicit expression. (This selectivity in attention does not indicate a decline in rigour, such as leaving a gap in a crucial place that left a whole range of cases unexplored.) For instance, the following would certainly be an acceptable gap in a journal article.

$$7\lambda + 6\mu = 186$$
$$17\lambda - 7\mu = 85$$

$$\Rightarrow \quad \lambda = 12, \quad \mu = 17$$

[13] Charles L. Fefferman, Arie Israel, and Garving K. Luli, "Sobolev Extension by Linear Operators" *Journal of the American Mathematical Society* 27 (2014): 69-70

In a piece of homework to be completed by a secondary school student, however, this 'gap' may contain the whole of the question to be answered, and so this would not be acceptable ('Show your working!').

Generally speaking, the extent to which explanatory gaps in the argument are acceptable increases with the mathematical sophistication of the readership, as they can fill more and more of the steps themselves. In fact, checking proof presentations is often a matter of having a general intuitive 'feel' for where the weak spots in an argument might be. These are then focused in on, and more standard parts may be more or less ignored. The mathematician William Thurston writes, 'I might look over several paragraphs or strings of equations and think to myself "Oh yeah, they're putting in enough rigmarole to carry such-and-such idea."' The existence of gaps is not problematic if the broader underlying logic of the argument is transparent: 'When the idea is clear, the formal setup is usually unnecessary and redundant – I often feel that I could write it out myself more easily than figuring out what the authors actually wrote.'[14]

Returning briefly to educational contexts; practicing writing in a way that gives explicit expression to every part of a proof is for obvious reasons a good idea if one is a novice at constructing rigorous arguments. So as noted, it is often a requirement for the first few example sheets in an introductory undergraduate analysis course, or a first course in axiomatic set theory. However, as the courses progress this insistence will typically be relaxed so as not to become tedious, with the implicit modal condition that the student could now fill in all the gaps in the arguments if necessary.

The fourth dimension of variability has already been discussed above: proof presentations will generally rely on many subsidiary results that are only mentioned and not proved explicitly. So usually only part of the entire proof of the theorem is required. In the context of a time-limited examination, when faced with a question such as 'Prove the Intermediate Value Theorem', students must often make a judicious guess as to how much they are supposed to derive from first principles and which subsidiary results they may simply quote. In a research context, however, mathematicians may rely on any result that has gained Public Acceptance. In the most common format, citations are indexed numerically with reference to a bibliography included at the end. The following passage from the same paper mathematical quoted above, which states results without including proofs, may be taken as typical:

> 'Let us compare our present results to what we know about $C^m(\mathbb{R}^n)$. Switching over to $\mathbb{X} = C^m(\mathbb{R}^n)$, we recall the following results [10, 12, 14].
>
> **Theorem 4**. *For any $E \subset \mathbb{R}^n$, there exists a linear map $T : \mathbb{X}(E) \to \mathbb{X}$ such that $Tf = f$ on E and $\|Tf\|_{\mathbb{X}} \leq C\|Tf\|_{\mathbb{X}(E)}$ for all $f \in \mathbb{X}(E)$.*
>
> *Moreover, if E is finite, then T has bounded depth.*

[14] William Thurston, "On Proof And Progress In Mathematics", *For the Learning of Mathematics* 15 (1995): 32

Theorem 5. *Let* $E \subset \mathbb{R}^n$ *be finite, and let* $N = \#(E)$.
Then there exists subsets $S_1, \dots, S_K \subset E$, *with* $K \leq CN$ *and with* $\#(S_k) \leq C$ *for each* k *such that*
$\|f\|_{\mathbb{X}(E)} \leq C \cdot \max_{1 \leq k \leq K} \left\|(f|_{S_k})\right\|_{\mathbb{X}(S_k)}$ *for all* $f \in \mathbb{X}(E)$.

Corollary 1. *Let* $E \subset \mathbb{R}^n$ *be finite, and let* $N = \#(E)$. *Then there exist linear functionals* $\xi_1, \dots, \xi_L : \mathbb{X}(E) \to \mathbb{R}$ *such that* $L \leq CN$, *each* ξ_ℓ *has bounded depth, and*
$c \cdot \max_{1 \leq \ell \leq L} |\xi_\ell(f)| \leq \|f\|_{\mathbb{X}(E)} \leq C \cdot \max_{1 \leq \ell \leq L} |\xi_\ell(f)|$ *for all* $f \in \mathbb{X}(E)$.'[15]

There are, of course, other restrictions on acceptability that have not been mentioned so far. Practical concerns may be in play: many academic journals have word limits for the articles they publish, for example. Journals will have stylistic guidelines that might extend to things like formatting. Certain 'styles of proof' are usual and expected for particular areas of mathematics such as group theory, or in analysis where the symbols ε and δ have an established usage. Insistence on the observation of these conventions enables arguments to be processed faster by the journal's readership. For more competitive journals, the conclusion itself will probably have to be interesting too.

In closing, we consider one other issue raised in the literature. David Tall distinguishes between three classes of arguments: those which are capable of convincing oneself, those which can be used to convince a friend, and those which can be used to convince an enemy.[16] We can expect some mathematicians to have a cautious professional scepticism towards the work of their peers. Yet it is also important to note that regarding someone as a professional colleague is again different from seeing him or her as an enemy. For example, if a gap in the argument occurs then they are unlikely to stubbornly insist that they do not know how to traverse it in cases where they could do so easily. And we can expect them to charitably invest a fair bit of time trying to understand those sections of a proof presentation where they suspect errors might have occurred.

1.v. Permanence, Reliability, Consensus and Autonomy

In this section, I will outline four core features that characterise contemporary mathematical practice. I will argue that these features are highly desirable, in that the working life of mathematicians would be much impoverished without them, and the progress of their enquiries would also be impaired. I will therefore refer to them as the 'Practical Virtues' of contemporary mathematical enquiry. In each case I will further argue for a second claim: that the rule of insistence on proof before publication is in an important way responsible for their ongoing maintenance. Taken together, these two claims show that mathematicians do have

[15] Charles L. Fefferman, Arie Israel, and Garving K. Luli, "Sobolev Extension by Linear Operators" *Journal of the American Mathematical Society* 27 (2014): 72.
[16] David Tall, "The Nature of Mathematical Proof", *Mathematics Teaching* 127 (1989): 30.

strong (though perhaps not sufficient) reasons for insisting upon proof for the Public Acceptance of results. The four Practical Virtues are as follows:

Permanence

When a statement becomes Publicly Accepted, it retains this status indefinitely.

Reliability

Publicly Accepted mathematical statements are always true.

Consensus

There is a shared agreement as to which statements are Publicly Accepted.

Autonomy

Competent mathematical researchers can always come to know Publicly Accepted results in an intellectually independent way, and mathematicians are never permitted to publish results on the basis of trust or authority alone.

Firstly, we discuss Permanence. In all of the natural sciences – even the most established of the physical sciences – the status of any hypothesis is always to some extent provisional. Researchers must always be open to the possibility that their most fundamental results will have to be revised in the light of new evidence. Historical examples abound: perhaps the most celebrated is Einstein's discovery of the merely approximate and local character of Newtonian mechanics, the most outstanding scientific achievement of its age, and long revered as a paradigmatic instance of the certainty that scientific work could aspire to. Some current scientific theories, such the central causal role of natural selection driving evolutionary change in biology, may seem so secure that the chances of their displacement are negligible. Yet we can always imagine future discoveries or experiments that might lead to this occurring.

The status of an established mathematical result, on the other hand, is not like this: results that have become Publicly Accepted are expected to remain part of mathematics on a permanent basis. Moreover, this expectation is not mere hubris, but is grounded in strong historical precedent. The theorems of Eudoxus and Archimedes are still our theorems, though the justifications for them may be quite different, and 'In most sciences one generation tears down what another has built, and what one has established, another undoes. In mathematics alone each generation adds a new storey to the old structure.'[17]

It seems plausible to think that this Permanence of Publicly Accepted mathematics might be a direct consequence of the second attribute: Reliability. That is, in actuality Publicly Accepted claims are rarely overturned simply because such results are rarely false: that established mathematics is, by and large, all true.

[17] Herman Hankel, *Die Entwicklung der Mathematik im letzten Jahrhundert*, 1884. Quoted in Lokenath Debnath, Dambaru Bhatta, *Integral Transforms and Their Applications* (London: Chapman and Hall/CRC, 2006), 315.

Taken together, Reliability and Permanence thus present an attractive picture of mathematical practice: only true results are Publically Accepted, and so over time only true – and hence incontrovertible – results are added to the structure.

We should, however, be careful about being so quick in inferring a high degree of Reliability from a high degree of Permanence. Although mathematical results that have been Publicly Accepted for a substantial period of time are almost never overturned, errors are actually quite often found in early drafts of arguments given by mathematicians, as well as in the years immediately following their publication. A number of examples are given in Section 4.iv. It is true that after surviving for a substantial period of time results are unlikely to later be found in error. Yet perhaps this is merely because such errors as might now remain must be unlikely to be revealed by the checking process, or even because interest in checking them simply subsides.

For mathematics that has been widely circulated, internalised and reconstituted, however, this suggestion gives expression only to a rather extreme form of scepticism. It is unreasonable to claim that there might still be errors in every single discourse purporting to present what we surely all believe is a genuine proof of Euclid showing that there are an infinity of primes. Too many thinkers have internalised the proof and come up with their own novel presentations. This is not mere rote checking, but deep and intuitive understanding. This also illustrates another topic of Chapter 4: how reliability is often only attained through a kind of social process.

So much for mathematics that has been sufficiently checked. However, another issue is that out of the huge number of new papers that are published each year,[18] a large proportion are not read by anyone other than the reviewer. Hence these results may not have been checked sufficiently thoroughly, and some of them may be untrue. If this is the case, large portions of the literature may not attain the high standards of Reliability characteristic of better-known mathematics.

To counter this suggestion, we need to say more about what is required for a piece of mathematics to count as Publicly Accepted. For although supplying a proof presentation that passes through the peer-review process is a necessary condition for Public Acceptance, it is not a sufficient one. Published results may enjoy varying degrees of centrality to mathematical research, and if a piece of mathematics meets with neglect and is relegated to obscurity, it would then no longer count as Publicly Accepted. On the other hand, if a result is included into the established body of widely circulated mathematics it will then be scrutinised more thoroughly. We may thus be confident that the central body of literature mathematicians rely on in producing new research is indeed highly Reliable.

Having given an overview of the Permanence and Reliability of mathematics, we now discuss the extent to which these two features are a consequence of the

[18] In the 1970's, Stanislaw Ulam estimated that around 100,000 theorems were proved each year, a figure later refined to nearer 200,000. See Stanislaw Ulam, *Adventures of a Mathematician* (Oakland: University of California Press, 1992), 288. This number would likely have now increased significantly.

insistence upon proof before publication. One observation which might make us doubt this connection is that although theorems themselves retain a high degree of Permanence, often arguments that are initially taken to support them must later be discarded – either because they are found to contain errors, or because they no longer meet increasingly demanding standards of rigour. This suggests that perhaps mathematicians' ability to decide if a theorem is true can outstrip their capacity for constructing and checking proof presentations. Indeed, full proofs were not available for many of the early calculus results discovered in the 18th century by Newton, Leibniz, Laplace and Lagrange and others – nor for much of Euler's work on infinite series – because the theoretical foundations underlying these areas of mathematics had not yet been fully worked out (see Section 5.v). Yet the results derived were, for the most part, true.[19]

One explanation of Euler's success is his use of non-deductive methods to corroborate his results. He knew he was proceeding on rather shaky theoretical ground, and so after performing his summations with algebraic manoeuvres he would check his answers with extensive numerical calculations.[20] Likewise, the new results of the calculus were often directed towards practical applications in physics and engineering, the success of which tended to increase confidence in them and provided good evidence that they were indeed correct.

However, in the next chapter we shall see that such non-deductive methods do not in general provide sufficient grounds for the Public Acceptance of results in mathematics. For instance, some results that seem empirically well supported, but which are denied Public Acceptance due to lack of a proof, later turn out to be false. Through exploring examples, it will become clear that proof does indeed play a central role in maintaining the high degree of Permanence and Reliability of mathematical literature.

Lastly, we consider why having a highly Reliable and Permanent literature is highly desirable and facilitates the progress of mathematical research. One benefit of Reliability concerns the dependence on previously established theorems. We have mentioned that mathematical arguments usually rely on subsidiary results in deriving their conclusions. Of course, both philosophers and scientists of all kinds also cite articles and books in support of their claims. But only mathematicians are warranted in the immediate acceptance of an auxiliary result just because a published article claiming it to be true is cited – even if they have not read the article themselves, have not heard of its author, or have not even looked up the name of the cited article in the bibliography. Without this reliance on quotation of subsidiary results research would be severely slowed down, and journal articles would multiply in length (the length of some journal articles is already presenting problems for researchers – see Section 1.vi).

A lack of Permanence would also impact upon this aspect of research in a related way. If results were accepted only for a time, it would be much harder to keep track of which articles were acceptable for a publishing mathematician to cite

[19] Judith Grabiner, "Is Mathematical Truth Time-Dependent?", *The American Mathematical Monthly* 81 (1974): 358

[20] Grabiner, "Is Mathematical Proof Time-Dependent?", 358.

when seeking to Publically Establish a new claim. The Permanence of mathematics is also part of what enables a 'deeper penetration into the subject-matter'[21] than in other fields of enquiry: because mathematical edifices are largely Permanent, over time they can develop to an immense complexity as every area of the subject-matter is scrupulously examined.

Lastly, the Reliability of mathematics means that natural scientists and others can take mathematical results 'off the shelf' without worrying about their veracity. They can regard the results they find in mathematics journals and textbooks as true *simpliciter*, and if their theories fail will rarely look to locate the error here.

We move now to our third Practical Virtue: that there is Consensus amongst mathematicians about which results have been Publicly Accepted and may now be stated as true without qualification. This third feature of practice is clearly connected to the previous two. It is in part because believing results proved in peer-reviewed mathematical journals is a Reliable source of gaining knowledge that such results are found convincing by all. Such Consensus, where rational, is merely the subjective acknowledgement of the objective phenomenon of Reliability. And if results moved in and out of Public Acceptance over time this would likely lead to widespread disagreements during the transitional periods.

To emphasise the high degree of Consensus within mathematics, consider first the range of disagreement amongst contemporary philosophers. In the analytic tradition today there are dualists, reductive and non-reductive materialists and idealists in the philosophy of mind; consequentialists, Kantians, contractarians, virtue and natural rights theorists, and non-cognitivists in ethics; formalists, constructivists, nominalists and Platonists in the philosophy of mathematics, and so on for each subdiscipline. Moreover, these disagreements constitute the permanent condition of academic philosophy.

This lack of Consensus does imply that philosophers lack the resources to discover true answers to the questions they pursue. However, it does suggest that philosophers find it difficult to formulate arguments that are able to establish lasting consensus amongst all of their academic colleagues about the truth and rational justifiability of their particular philosophical position.[22] This remains the case even though consensus may be achieved in a negative direction; Gettier's paper on the justified true belief account of knowledge is perhaps a notable example of this.[23]

In mathematics, however, such disagreement is the exception rather than the rule. Consider the difference between an undergraduate course in moral philosophy on the one hand, where one can expect to learn about key historical figures such as Mill, Bentham, Kant, Aquinas and Aristotle, the different conceptions of morality

[21] Arthur Jaffe and Frank Quinn, ""Theoretical Mathematics": Toward A Cultural Synthesis Of Mathematics And Theoretical Physics", *Bulletin of the American Mathematical Society* 29 (1993): 2

[22] Alasdair Macintyre, "On Having Survived the Academic Moral Philosophy of the Twentieth Century", in *What Happened in and to Moral Philosophy in the Twentieth Century*, ed. Fran O'Rourke (Notre Dame: Notre Dame University Press, 2013), 18.

[23] Edmund Gettier, "Is Justified True Belief Knowledge?", *Analysis* 23 (1963): 121-123.

they have each articulated, and the arguments supporting their positions; and an undergraduate course in mathematical analysis on the other, where one can expect a concise and ahistorical presentation of what everyone agrees are the established theorems and true results pertaining to the central questions in that area. Moreover, unlike in natural science where consensus is perhaps maintained only temporarily by a group of researchers operating under a shared paradigm in the course of what Thomas Kuhn has called 'normal science', in mathematics the agreement reached is Permanent.[24]

We now discuss whether insistence on proof does play a central part in maintaining the Consensus just described. Historical evidence suggests it does: in periods where proof has not been available, such as the early development of the calculus in the 17th and 18th century, Consensus has often been lacking. Let us consider the matter further. What is at issue is whether the policy of proof prior to publication will lead to mathematicians agreeing that published results are categorically true. Clearly, if a mathematician agrees that a proof has been given, they must join the Consensus: deductive reasoning always rationally compels assent. So the question reduces to whether mathematicians do in general believe that the discourses published in journal articles are successful proof presentations.

Generally speaking, mathematicians reading a published argument will agree that it gives expression to a proof. If there are serious doubts about the adequacy of a proof presentation, then a retraction of the article may be recommended, and a more satisfactory presentation sought. This rarely happens, however. Recall that one necessary criterion for a discourse to count as an acceptable proof presentation is that the peer reviewer judges that it will be found convincing by the entire readership of that journal. Moreover, mathematicians can operate effectively whilst only publishing articles meeting this criterion – one that would be far too restrictive and demanding for other disciplines such as philosophy. If the proof presentations published in a particular journal were often found unconvincing by mathematicians, action would be taken to rectify the situation to keep the reputation of the journal intact.

This agreement is also aided by the standardisation of proof presentations mentioned in Section 1.i. There are accepted classes of manoeuvres that will not be questioned, and standard lines of attack that all working mathematicians have been trained to use. This further reduces the chances of disagreement.

Let us now consider the attitude of the entire mathematical community to a published result, as opposed to just those that have read the article. It not true that the mere existence of a proof presentation that has passed the peer-review checks compels assent to a result on pains of brute irrationality: we have already noted that sometimes such discourses contain mistakes. However, for the most part, we can expect widespread agreement to occur, given that mathematicians constitute a community of professionals pursuing shared goals (recall here Tall's distinction from Section 1.iii).

[24] Thomas Kuhn, *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1996), 10.

The benefits of Consensus amongst mathematicians are again clear. As with Reliability, it is important for the practice of citing other articles when invoking subsidiary results. If the auxiliary results needed were not agreed to be Publically Accepted by all researchers, then the effectiveness of collaborative mathematical research would be reduced here. It is not sufficient that the literature is highly Reliable and cited results are always in fact all true: this also needs to be acknowledged in practice.

The fact that there is a more or less permanent Consensus about answers to fundamental questions and the truth of basic results also means that mathematicians are free to pursue intensively specialised research and are not forced to spend time and energy warring with competing schools within the discipline. This is again similar to Kuhn's account of the necessary conditions for normal science to progress, although as we have already said in mathematics such agreement is Permanent, and not periodically interrupted by crises wherein it is subject to fragmentation.[25]

Before moving on to the fourth Practical Virtue of mathematical practice, a caveat must be added. Throughout history there *have* been disagreements of sorts between competing schools within mathematics: alternative paradigms and approaches; questions of how best to formulate concepts, which theorems are interesting, which objects or problems are the most important to study; aesthetic judgements and questions of explanatory value, and so on. Consider also the rival attempts to rigorise the calculus: geometric, algebraic, arithmetic (again, see Section 5.v). However, apart from occasional exceptions, such as the disagreements with classical mathematics of intuitionists led by Brouwer about what constitutes an acceptable mathematical argument, which were in any case fringe, these divergences all occur prior to the stage at which the question of finding a proof arises.

What is always agreed upon is that the conclusion is indeed a consequence of the premisses and rules of inference employed. For instance, rival mathematical systems such as alternative formulations of geometry are not incompatible in the way that rival philosophical theories concerning which set of actions are morally justified or rival scientific hypotheses about the nature or cause of a particular physical phenomenon are. They merely signify that a change in subject matter has occurred, and hence do not express the kind of disagreement amongst practitioners relevant to Consensus.[26] There are of course disagreements about the best or most fruitful way to proceed, but generally speaking once a set of starting points and tools are agreed upon – and in contemporary mathematics it is usually possible to be clear what these are – disputes between contending parties may be resolved by seeing which of them can find a proof for their claims.

The fourth and final Practical Virtue I will draw attention to is Autonomy. This feature of mathematical practice has two components. Firstly, any competent mathematical researcher can in principle come to find their own explicit reasons for believing any Publicly Accepted claim simply by reviewing the relevant

[25] Kuhn, *The Structure of Scientific Revolutions,* 66.
[26] Ernest Cassirer, *The Problem of Knowledge* (Connecticut: Yale University Press, 1978), 32.

literature – for example, the article in which it was first announced as true. Secondly, no mathematician is ever permitted to publish a result on the basis of their personal authority alone. When journal referees judge that a printed argument constitutes a sufficient basis for publication, they never rely upon trust in the testimony of the publishing mathematician. The Consensus just described is thus spontaneous rather than coerced; this is what the philosopher Jody Azzouni has called 'The benign fixation of mathematical practice'.[27]

A clarification must be added here, stressing the partly modal and partly social nature of this Practical Virtue. It may be the case that no mathematician has personally scrutinised proof presentations for all or even a large proportion of the mathematical results they believe and regard as Publicly Accepted: this is not what is intended. Rather, what is important is that any competent mathematician could in principle find their own reasons to support any one of their Publicly Accepted mathematical beliefs simply by referring to a suitable place in the existing literature, and that Publicly Accepted results are supported by explicit argumentation that has been thoroughly checked by at least a sizable number of mathematicians, so that results are never put forth merely on the basis of trust or authority alone.

Insisting on proof prior to publication of results is clearly an effective way of maintaining Autonomy because reading such a discourse enables any competent reader to know its conclusion in an intellectually independent way. ('Competence' here again includes being able to fill in the gaps in the proof presentation, which may require substantial work.) It is also hard to see how Autonomy could be adequately maintained without this restriction: if an argument is supported merely by inductive evidence rather than proof, we will be required to accept the judgement of the author that the evidence is conclusive in this instance (see Section 2.v for further explanation of this point).

Autonomy is valuable for its own sake, and also for maintaining the other three Practical Virtues. For instance, authority is often less effective in maintaining lasting Consensus. Russell writes, 'Have no respect for the authority of others, for there are always contrary authorities to be found.'[28] In contrast to this, whilst a student is still learning to think mathematically, or to understand the use of some new concept, some reliance on the authority of a teacher may be necessary. Results or techniques must be presented in some particular order, and the justification of a principle might take a lesson too far afield at the stage when it is first needed. But at some point in their development towards becoming an independent mathematician, students usually begin to insist on their own reasons for believing results they are taught.

[27] Jody Azzouni, "How and Why Mathematics is Unique as a Social Practice", in *18 Unconventional Essays on the Nature of Mathematics*, ed. Reuben Hersh (New York: Springer, 2006), 208.
[28] Bertrand Russell, "A Liberal Decalogue", in *Autobiography* (London: Routledge, 2009), 534.

1.vi. Recent Developments in Mathematics

In this section I will discuss two recent developments in mathematical practice: that proof presentations for Publicly Accepted results now sometimes run to exorbitant lengths, and that in the discovery of proofs for Publicly Accepted results mathematicians are sometimes now permitted to make essential usage of computers to check though the argumentation, in such a way that no explicit, surveyable written proof presentation is produced. Through discussing a pair of case studies for each, I shall argue that in these contexts the effectiveness of proof in establishing and maintaining the Practical Virtues is significantly impaired.

First we discuss how the proof presentations mathematicians give are now sometimes extremely long, which threatens to thwart the checking process. Modern mathematical research has now reached a point of incredible depth and complexity, and as deeper and more complex results have been discovered, increasingly long proof presentations have appeared in the literature in recent decades. Some examples follow:

- In 1995, Andrew Wiles' celebrated proof of Fermat's Last Theorem was published in a paper over 100 pages long.[29]

- Between 1983 and 2004, Neil Robertson and Paul Seymour published a proof of the Robertson-Seymour Theorem over the course of some twenty papers. Also known as the Graph Minor Theorem, this result says that in any infinite collection of graphs there is a pair such that one is a minor of the other. The proof presentation totaled over 500 pages.[30]

- In the 1980's, a proof of the Almgren Regularity Theorem in geometric measure theory was given by Frederick Almgren. His manuscript totaled 1728 pages and was later published in an edited form in a book of nearly 1,000 pages.[31]

- In the year 2000, Laurent Lafforgue proved Lafforgue's Theorem, part of a series of conjectures in various parts of algebra known as the 'Langlands Program'. The proof was published in three papers totaling around 600 pages. Lafforgue received a Fields Medal for his efforts.[32]

- In 2006, a proof of the Strong Perfect Graph Theorem appeared in the *Annals of Mathematics*. Published by Maria Chudnovsky, Neil Robertson, Paul Seymour and Robin Thomas, the presentation totaled nearly 180 pages.[33]

---

[29] Andrew Wiles, "Modular Elliptic Curves and Fermat's Last Theorem", *Annals of Mathematics* 142 (1995): 443-551.
[30] Reinhard Deistel, *Graph Theory* (Berlin: Springer-Verlag, 2010), 333.
[31]Frederick Almgren, *Almgren's Big Regularity Paper,* ed. Vladimir Scheffer and Jean Taylor (Singapore: World Scientific Publishing, 2000).
[32] Gérard Laumon, "The Work of Laurent Lafforgue", *International Congress of Mathematicians* 1 (2002): 94.
[33] Maria Chudnovsky et. al., "The Strong Perfect Graph Theorem", Annals of Mathematics 165 (2006): 51-229.

Shortly we shall discuss a proof presentation whose length is far greater still; before that we turn to our first of four case studies.

In a 1979 paper, De Millo, Lipton and Perlis discuss what was then a recent example of the effects of the growing length and complexity of proof presentations. Two sets of researchers – one American, one Japanese – had independently made claims regarding the same homotopy group, for which they each thought they had discovered proofs. However, the two claims contradicted each other. So at least one group of researchers must have been mistaken. But because of the length and complexity of the arguments offered by both groups of researchers, it was not obvious where the mistake was. So, the research groups exchanged publications, and each tried to find an error in the others' work. Yet despite the obvious motivation for finding a mistake, neither team of researchers was able to discredit the argument given by the other. A third group later presented an argument supporting the American team, leading the Japanese mathematicians to temporarily withdraw to check over their research.[34]

In this instance, then, the fact that proof was insisted upon was not enough to establish Consensus between the two groups of researchers: the discourses they supplied were of such a length that they could both be plausibly put forth as proof presentations, and mathematicians were for a long time unable to distinguish which if either of these claims was correct. Moreover, it also illustrates a suggestion made earlier: perhaps the degree of Reliability of accepted results – and especially of results sustained only by very long proofs such as these – is actually lower than we may think, because errors are often very difficult to detect – even though for longer arguments one would think the occurrence of at least one minor error would become increasingly likely.

Our second case study is another theorem from the literature that has generated debate about mathematical practice. This concerns the classification of finite simple groups[35] – the so-called 'Enormous Theorem'. Finite simple groups are in a sense the building blocks for all groups, in the same way that all natural numbers can be decomposed into their prime factors, although for groups the decomposition series is not necessarily unique.[36] The Enormous Theorem states that every finite simple group belongs to one of the following four categories: a cyclic group of prime order, an alternating group of degree at least 5, a group of lie type, or one of 26 'sporadic' groups which do not fit into these three classes.

The majority of the proof of this theorem was discovered in the years 1950-1980, leading to Daniel Gorenstein – the chief architect of the proof – describing it as the 'thirty years war'.[37] However, some parts of the argument date back as far as

[34] Richard De Millo, Richard Lipton and Alan Perlis, "Social Processes and Proofs of Theorems and Programs", *Communications of the ACM* 22 (1979): 272. Sadly, the authors do not supply a reference, so the eventual outcome of the debate is unclear.

[35] A group is simple iff it has no non-trivial normal subgroups. A subgroup H of a group G is normal iff it is invariant under conjugation: $\forall g, h \in \mathrm{G}, \quad h \in \mathrm{H} \Leftrightarrow ghg^{-1} \in \mathrm{H}$.

[36] The Jordan-Hölder Theorem shows it is unique up to permutation and isomorphism, however.

[37] Ronald Solomon, "On Finite Simple Groups and Their Classification", *Notices of the AMS* 42 (1995): 231.

1899.[38] Under Gorenstein's leadership, the effort required contributions from over 100 mathematicians from many different countries around the world. Gorenstein announced discovery of a complete proof in 1983, and began publication in 1994 with Lyons and Solomon.[39] Actually, at this time part of the proof for quasithin groups was not yet complete, and was only dealt with in 2004 by Aschbacher and Smith.[40] Though this was thought to be the last gap in the argument, a second minor gap emerged in 2008 and was corrected by Solomon and Harada.[41]

The proof attracted interest even prior to its publication. Despite the daunting 200-page chunks in which the preprints were issued, many individuals and seminar groups persevered with reading the parts of the proof that were presented up to 1975. Over the next five years, however, at least 3,000 compact pages of mathematical argumentation were circulated, which 'simply overwhelmed the digestive system of the group theory community.'[42] The length of the final proof presentation was staggering – Gorenstein estimated it to be around 15,000 pages. A shorter second-generation proof presentation has since been under construction – though the expected length is still around 5,000 pages. One reason for this exorbitant length is that the classification itself is so complex.[43]

The Enormous Theorem is important for group theorists because it enables 'divide and conquer' strategies for proving general results about finite simple groups: mathematicians may simply show that such a claim holds for finite simple groups of each kind. It has therefore been cited in further journal articles and is clearly Publicly Accepted by mathematicians. However, the modal 'in principle' qualification of Autonomy must be stretched to breaking point here: the average mathematician simply cannot be expected to work through even a 5,000-page second-generation presentation of the proof. If mathematicians' hard-won Consensus is not to fragment, then most practitioners must simply believe the result on the collective testimony of those group-theorists who have worked on it, and of those enthusiasts who have been willing to invest substantial portions of their lives checking through it.

In addition to a loss of Autonomy, there are clearly also questions about the Reliability of the result. Solomon writes 'Is the database correct? Is there a 27th

---

[38] Ronald Solomon, "A Brief History of the Classification of Finite Simple Groups", *Bulletin of the American Mathematical Society* 22 (2001): 315.

[39] Daniel Gorensten, Richard Lyons and Ronald Solomon, "The Classification of the Finite Simple Groups, Part 1", *American Mathematical Society Surveys and Monographs* 40.1 (1994).

[40] Michael Aschbacher and Stephen Smith, "The Classification of Quasithin Groups: I. Structure of Strongly Quasithin κ-groups", *Mathematical Surveys and Monographs* 111 (2004). Michael Aschbacher and Stephen Smith, "The Classification of Quasithin Groups: II. Main Theorems: The Classification of Simple QTKE-groups", *Mathematical Surveys and Monographs* 112 (2004).

[41] Koichiro Harada and Ronald Solomon, "Finite groups having a standard component L of type $\widehat{M}_{12}$ or $\widehat{M}_{22}$", *Journal of Algebra* 319 (2008), 621–628

[42] Ronald Solomon, "On Finite Simple Groups and Their Classification", 236.

[43] The sporadic groups alone comprise 5 Mathieu groups, the 3 Conway groups, the 3 Fischer groups, the Higman-Sims group, the McLaughlin group, the Held group, the Rudvalis group, the Suzuki sporadic group, the O'Nan group, the Harada-Norton group, the Lyons group, the Thompson group, the 4 Janko groups – the fourth of which was the last to be discovered, in 1976 – the baby monster group, and the Fischer-Griess Monster group, which has an order of 808017424794512875886459904961710757005754368000000000.

sporadic simple group? I seriously doubt it, but it would be chutzpahdich to assert that a 5000-page 40-year human endeavor is beyond the possibility of human error.'[44] Indeed, mathematicians' confidence in the result is partly based on the fact that no new groups have been discovered in recent years after the initial errors mentioned earlier came to light. For the average mathematician, this is a move away from explicit, deductive reasons for belief and towards relying on testimony and non-deductive evidence.

We move on to a second pair of case studies, now concerning computer-based mathematics. Over the last few decades, many important results have become Publicly Accepted on the basis of arguments that make essential use of computers. These computers may be used merely to check through large amounts of computations, but they may also have a more involved role and even construct substantial sections of the reasoning behind the argument. Results whose justifications are of this nature that have been widely reported in the public domain include the verification that games of draughts will be drawn if both players play optimally,[45] that a Rubik's Cube can always be solved in 20 moves or less,[46] and that the solution to a Sudoku puzzle needs at least 17 clues.[47] We now discuss the proof of a more weighty mathematical result in detail.

In the 1850's, Francis Guthrie – who was then a student at University College London – discovered a way of colouring the map of England's counties using four colours such that no two counties sharing a common border receive the same colour. Consider a general map divided into regions. Do four colours always suffice to achieve this result? In essence, the claim that they do is our third case study: the famous Four Colour Theorem. We need only replace the intuitive idea of a map with the more precise concept of the infinite Euclidean plane containing a planar graph drawing – that is, a finite number of points or vertices joined by non-intersecting curves of finite length with endpoints at the vertices,[48] which thus divide the plane into one infinite region and a finite number of finite regions. After making no progress with the general result, Francis discussed it with his brother Frederick, who shared it with his teacher, the eminent mathematician Augustus De Morgan. The problem was later studied by other mathematicians including Cayley and Minkowski. In 1879, Kempe gave an argument purporting to establish the theorem, but a flaw was found after 11 years by Heawood in 1890.[49]

The essence of Kempe's argument was as follows. Firstly, given any graph drawing in the plane, we construct a dual graph drawing by placing a vertex on the interior of every region and joining vertices contained in two regions iff those regions share a common border, as shown in the diagram below.

---

[44] Solomon, "A Brief History of the Classification of Finite Simple Groups", 347.
[45] Justin Mullins, "Checkers 'solved' after years of number crunching", *New Scientist*, 19th July 2007. The argument is due to Jonathan Schaeffer, who had been working on the problem for 18 years.
[46] Tomas Rokicki et. al., "The Diameter of the Rubik's Cube Group is Twenty", *Siam Journal on Discrete Mathematics* 27 (2013): 1082-1105.
[47] "Mathematicians Solve Minimum Sudoku Problem", *MIT Technology Review*, 6th January 2012. The proof was found by Gary McGuire, Bastian Tugemann and Gilles Civario.
[48] Farey's theorem tells us that we can replace this phrase with 'straight line segments'.
[49] John Mitchem, "On the History and Solution of the Four-Color Map Problem", *The College Mathematics Journal* 12 (1981), 108-116.

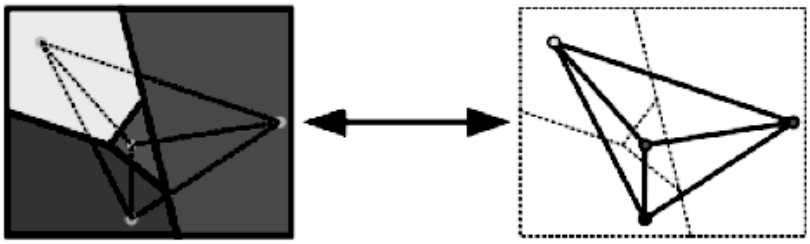

Figure 1.7. The process of constructing a dual graph

This reduces the problem to that of colouring the vertices of a given planar graph drawing with four colours so that no two adjacent (i.e. connected with an edge) vertices receive the same colour. For technical reasons it is also useful to 'triangulate' the graph drawing by adding more edges until the boundary of each region has exactly three edges. This move is clearly permissible because it serves only to restrict the possible colourings.

Kempe's argument now proceeds as follows. Firstly he proved that any triangulated planar graph with at least 5 vertices must have either a vertex[50] of degree 3, 4 or 5. Then he argued by mathematical induction. Suppose there are graphs that cannot be four-coloured. Let G be one of these graphs such that the number of vertices of G is as small as possible. Clearly, if such a graph exists it must have at least 5 vertices, and so will have a vertex of degree 3, 4 or 5. Suppose G has a vertex v of degree 3. If we remove this vertex and the three edges incident to it, the resulting graph will have fewer vertices than G, so by the definition of G it must be four-colourable. But when we add v back in, only at most three of the four possible colours will be used up by its neighbours, leaving one colour available for v (see Diagrams 1.8-1.10 below).

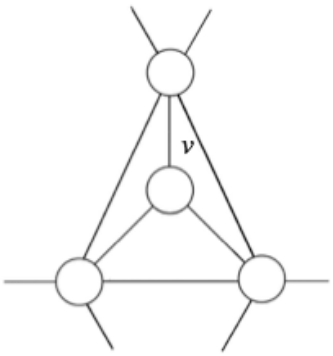

Diagram 1.8. A vertex of degree 3.

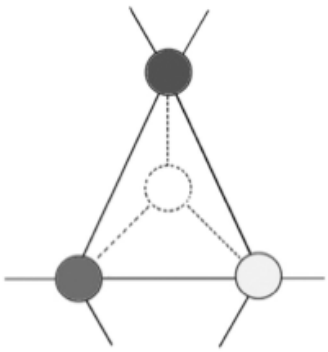

Diagram 1.9. Remove v and colour the remaining graph.

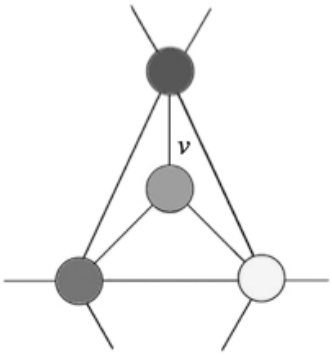

Diagram 1.10. Extend this to a four-colouring of T.

[50] The degree of a vertex is defined as the number of edges incident to it.

Kempe tried to show that similar procedures – although rather more complex, and involving technical devices known as 'Kempe Chains' – could be employed for vertices of degrees 4 and 5. Because every graph must contain a vertex of one of these degrees, then we can employ at least one of the three sub-arguments to show that G is four-colourable after all, which is a contradiction. So in fact no such graph G can exist, and all graphs are four-colourable. But as already mentioned, there was a flaw in his argument: in 1890, Heawood produced a graph for which Kempe's procedure for a vertex of degree 5 did not work, although the graph itself could easily be four-coloured by other means.[51]

Almost a century later in 1976, a proof of the theorem was finally discovered by Haken, Appel and Koch.[52] The logical structure of their proof is similar to Kempe's argument, using minimal counterexamples and extending colourings of the graph drawings produced when parts of them are removed, or replaced with smaller structures that have fewer vertices, leaving fewer vertices overall. Though this is a slight oversimplification, for our purposes, we can imagine the structure that is replaced as comprising the interior of a cycle or 'ring', as in Diagrams 1.8-1.10 above. A structure to be deleted is called a 'primary configuration', and that which replaces it is called a 'secondary configuration'. When a configuration H comprises the total interior of a ring in G in this way, we say that G 'contains' H.[53] A primary configuration H is said to be 'reducible' if for any four-colouring of G with a ring containing just H it is always possible to extend a modified version of G, with H removed and some particular secondary configuration that has fewer vertices than H added in its place, to a four-colouring of G itself.

Lastly, to complete the proof we need to make use of the concept of unavoidability. A set S is unavoidable if every graph G with at least 5 vertices must contain some member of S. Hence, we need only find an unavoidable reducible set to prove the theorem. Unfortunately, the smallest such set the three mathematicians could find had nearly 1900 elements.[54] Moreover, some of these configurations were themselves rather complicated, especially compared with the simple one-vertex subgraphs Kempe had tried to make do with. Yet even in Kempe's attempted proof presentation the arguments for reducibility were quite complex – indeed, enough so to enable his mistake to go unnoticed for eleven years. The argumentation needed to complete the final proof was so complicated as to elude construction by hand: it required around 1200 hours of computation time on what was then a high-speed computer.[55] Moreover, the computer was not just performing simple checks but had to construct many fairly complicated

[51] Percy Headwood, "Map-Colour Theorem", *Quarterly Journal of Mathematics,* 24 (1890): 332–338

[52] Kenneth Appel and Wolfgang Haken, "Every Planar Map is Four Colourable, Part I: Discharging", *Illinois Journal of Mathematics* 21 (1977): 429-490. Kenneth Appel, Wolfgang Haken and John Koch, "Every Planar Map is Four Colourable, Part II: Reducibility", *Illinois Journal of Mathematics* 21 (1977): 491-567.

[53] This implies but is not equivalent to the subgraph relation; a number of authors seem to conflate these two concepts.

[54] Frank Bernhart, "A Digest of the Four Colour Theorem", *Journal of Graph Theory* 1 (1977): 207.

[55] Bernhart, "A Digest of the Four Colour Theorem", 207.

individual arguments, each analogous to the one given above for a vertex of degree three, in order to show reducibility in each case.

The essential use of computers in finding the proof caused much debate, but the result gradually gained Public Acceptance by mathematicians. Unlike the Group Classification Theorem, there seems now to be no question about its truth. William Thurston writes, 'I interpret the controversy as having little to do with doubt people had as to the veracity of the theorem or the correctness of the proof',[56] and indeed as shall see in Section 5.vi the proof has now been formalized and verified by Georges Gonthier.[57] Yet the example again serves to undermine Autonomy, because individual mathematicians are not able to come to believe the four-colour theorem for their own direct reasons that are understood in detail, but rather only because of the 'testimony' of a computer. This has led some philosophers such as Thomas Tymoczko to conclude that mathematicians' knowledge of the Four-Colour Theorem should not be considered *a priori*, as it is based only the outcome of a kind of empirical computer-based experiment.[58]

Our fourth and final case study is Thomas Hales' proof of the Kepler Conjecture. This is the claim that of all the possible methods of packing identical spheres into space, the one shown below is the most efficient (as intuition might suggest).[59]

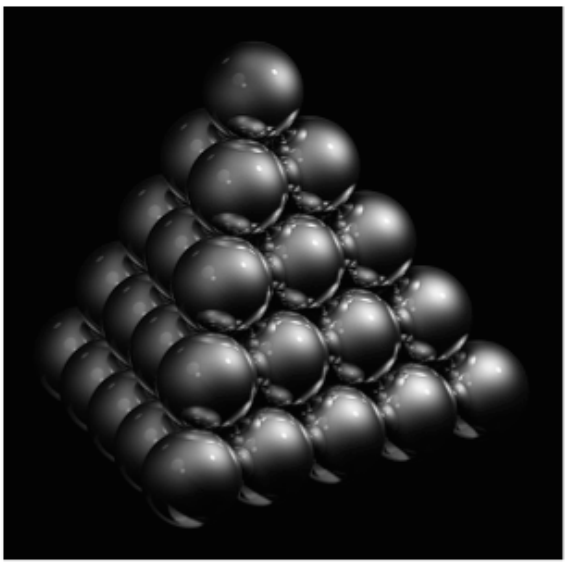

Figure 1.11. The most efficient way of packing spheres.
http://commons.wikimedia.org/wiki/File:
Close-packed_spheres.jpg

[56] William Thurston, "On Proof and Progress in Mathematics", *For the Learning of Mathematics* 15 (1995): 29.
[57] The meaning and implications of this claim will also be made clear in Section 5.vi.
[58] Thomas Tymoczko, "The Four-Colour Problem and its Philosophical Significance", *The Journal of Philosophy* 76 (1979): 57-83.
[59] Thomas Hales, "A Proof of the Kepler Conjecture", *Annals of Mathematics* 162 (2005): 1065-1185.

Hales submitted his argument – which again relied heavily on computers – to the prestigious journal *Annals of Mathematics* mentioned above, where his case was considered for three years. Eventually, the journal came to a decision that has been integrated into an explicit policy relating to cases like this, and which now features on their website. The editors state that arguments which make essential use of computers will be considered for publication only for 'exceptionally important mathematical theorems', wherein the Annals may publish the 'human' part of the proof, which reduces the problem to one solvable by a computer program, as well as storing the computer code and supplementary documentation on its website. They also write that 'The computer part may not be checked line-by-line, but will be examined for the methods by which the authors have eliminated or minimized possible sources of error'.[60]

Similarly to the previous case study, and to all other cases where results are only established in a way that relies on automated computer-based methods, the reasons for belief available to individual mathematicians for believing the Kepler Conjecture cannot be characterised at a detailed mathematical level, but only in terms of the properties and capabilities of certain computer systems and why the outputs they have given reveals that *some* such detailed set of mathematical reasons for belief must be available. Clearly, this is an undermining of the Autonomy enabled by more traditional methods.

1.vii. Conclusion

In this chapter we have seen how contemporary mathematics incorporates a rule whereby mathematicians are always required to give a proof presentation in order to unqualifiedly assert as true new results in serious mathematical publications – and hence that proof is necessary for the Public Acceptance of results. The rational justification of this rule – whether mathematicians should continue to submit to it – will be the central question discussed in this thesis.

Adherence to this rule is centrally important for maintaining the four Practical Virtues of mathematical practice: Permanence, Reliability, Consensus and Autonomy. Moreover, these are important for the continuing progress of mathematical enquiry itself. This suggests that a partial rational justification for the rule is available at this stage. Yet we have also seen that where new results are established through proof presentations that are excessively long, or rely essentially on computers, proof furnishes only a less effective means to securing them. So despite the continued insistence on proof, the Practical Virtues are in danger of declining in future.

It was also pointed out that in earlier periods when deductive techniques could not be fully relied upon mathematicians such as Euler used non-deductive methods as effective aids to their research. The next chapter, then, we present an overview of non-deductive methods in mathematics, together with an assessment of their suitability for justification in the context of both Public and Private acceptance.

---

[60] "Statement by the Editors on Computer Assisted Proof", accessed 23rd June 2015, http://annals.math.princeton.edu/board

2. Non-deductive Arguments

This chapter provides a general discussion of non-deductive methods and their role within mathematical practice. Perhaps their most important use lies in the discovery of plausible conjectures that are later proved, and as a check on our deductive work. Yet the use of non-deductive methods may also leave us strongly inclined to yield Private Acceptance to a conjecture prior to a proof having been found, especially when computers have been used to generate extensive data. It also appears that mathematicians themselves have come to Privately Accept claims such as the Goldbach Conjecture on these kinds of grounds alone. However, I will argue that with the possible exception of certain probabilistic methods – whose reliability can be explicitly evaluated – non-deductive arguments are unsuitable for justification in the context of Public Acceptance. Their use here would lead to the further deterioration of the four Practical Virtues given in Chapter 1: Permanence, Reliability, Consensus, and Autonomy.

2.i Non-deductive Methods

A non-deductive method is a mathematical technique that does not yield a proof, though it may have some other function: for example, it could be a source of plausible new conjectures. Non-deductive methods may be organised into categories: inductive, experimental, visual, analogical. Rather than giving a complete definition of non-deductive methods, I will instead briefly discuss and give examples of each of these four kinds of non-deductive method in turn. The list is not intended to be exhaustive, and the categories may overlap.

One example of an inductive procedure is to take a general claim we think is plausible and then check a large number of cases. For example in number theory, if we have a conjecture we think holds true for all natural numbers, we can check a large number of them to verify it these particular instances. Mathematicians sometimes use the term 'empirical evidence' to describe such a collection of data, in contrast to its usual philosophical use in connection with sensory experience, although here the adjective 'quasi-empirical' will be preferred.

In the previous chapter, I mentioned that a vast amount of this kind of inductive work has been carried out in the case of the Goldbach Conjecture. This will be discussed in much greater detail in Section 2.v below. A different use of this kind of technique was exemplified in 1995 with the discovery of the Bailey-Borwein-Plouffe formula, which gives an exact expression for the n^{th} digit of π in base 16 without calculating any other digits. The formula is as follows:

$$\pi = \sum_{i=0}^{\infty} \frac{1}{16^i}\left(\frac{4}{8i+1} - \frac{2}{8i+4} - \frac{1}{8i+5} - \frac{1}{8i+6}\right)$$

This advance was only made possible by extensive searching using a procedure called the PSLQ integer relation algorithm, although once the formula had been discovered the proof was fairly straightforward.[61]

Experimental reasoning in the narrow, literal sense means to employ a physical simulation to explore some area of pure mathematics. For instance, the 19th Century Belgian physicist Joseph Plateau made extensive studies with soap films stretched over wire frames, exploiting their material properties to derive results in geometry concerning minimal surfaces under given boundary conditions. The construction of such a surface is now known as a Plateau Problem, and further progress has been made in this direction since.[62] Consider another, related example: when a soap film forms a closed surface in space with no holes, enclosing a fixed volume of air, it soon takes the familiar shape of an approximately spherical bubble. Because this shape is the outcome of surface tension and pressure forces acting on the film, this suggests that the sphere is the figure of given surface area that encloses the largest volume in space (a result proved by Hermann Schwarz in 1884).[63]

Another example of an experimental technique in being deployed in mathematics is Leonard Adleman's algorithm in graph theory, which uses the physical properties of DNA molecules to find Hamiltonian paths in a given graph. This will be discussed in detail in Section 6.i. A third example gives a solution to the problem in graph theory of finding the shortest path(s) between two vertices. We use rings to represent each vertex and connect all adjacent vertices with pieces of string of the same length. If we then pull apart the two rings corresponding to the points we are interested in, the taut strings show us where the shortest path lies.

Experimental reasoning in this literal sense is now fairly rare in mathematics: journals like *Experimental Mathematics* use the term only to indicate that non-deductive evidence is playing a merely analogous role to experimentation within the nature sciences. Yet many results in Euclidean geometry were first discovered in the context of practical mensuration problems that arise in astronomy, navigation, farming, surveying and commerce. Indeed, traditionally the origin of geometry – literally, 'land-measurement' – is said to lie in Ancient Egyptian attempts to ensure a fair reallocation of land amongst the people after the annual flooding of the Nile river had washed away all recognisable landmarks.[64]

One use of the human visual system in mathematics is to draw an accurate diagram of a given geometrical setup and observe that two quantities visually appear to bear some relation to one another. For instance, two angles might appear equal in size, or two line segments of equal length. Once this has been noticed a

---

[61] David Bailey, Peter Borwein and Simon Plouffe, "On the Rapid Computation of Various Polylogarithmic Constants", *Mathematics of Computation* 66 (1997): 903-913.
[62] Jenny Harrison, "Soap Film Solutions to Plateau's Problem", *Journal of Geometric Analysis*", 24 (2014): 271-297.
[63] Frank Morgan, "Soap Bubble Clusters", in *Expeditions in Mathematics*, ed. Tatiana Shubin, David Hayes and Gerald Alexanderson (Washington: Mathematical Associate of America, 2011), 165.
[64] Proclus, *A Commentary on the First Book of Euclid's Elements*, trans. Glenn Morrow (Princeton: Princeton University Press, 1992), 52.

deductive proof of the relation can then be sought. Diagrams can also play various other roles in finished geometrical discourses, such as providing a convenient psychological prop to the reader, or even featuring essentially in the exposition of a proof itself.[65] A classic example of the use of diagrams is Socrates' conversation with a slave boy in the *Meno*. We shall also see in Section 2.iii that the scope of diagrammatic reasoning in geometry has also been enhanced by technological developments in recent years.

One example of reasoning by analogy might be to appeal to a similarity that is exploited in constructing new conjectures. For instance, we can often usefully compare geometrical objects we want to know about to their analogical equivalents in other dimensions: the sphere to the circle; the cube to the square. We said above that of all solids with fixed surface area the sphere encloses the most volume in space. This makes it reasonable to investigate the claim that of all the closed curves of fixed length in the plane the circle encloses the most area. Finding such a curve is known as the 'isoperimetric problem', and has been discussed since antiquity. We can also fruitfully generalise the problem to n dimensions.

Another example of analogy is when mathematicians notice a broad similarity of structure between two apparently very different domains, such as the ring of linear maps from an n-dimensional vector space V over a field $\mathbb{F}$ to itself, with binary ring operations given by the composition and pointwise addition of these maps, and the ring of $n \times n$ matrices with entries from $\mathbb{F}$ under matrix multiplication and addition. After establishing this connection explicitly by giving an isomorphism, we can prove results about linear maps by looking at corresponding results about matrices (or conversely). But even prior to this kind of explicit theoretical work being carried out, such a connection may suggest new conjectures that can be proved on an individual basis.

Lastly, a final example of reasoning by analogy is provided by the *fragment on lunes* of Hippocrates of Chios, the earliest substantial mathematical text from Ancient Greece whose origin has been ascertained with certainty.[66] A lune is a figure formed by the intersection of circular arcs, with one side convex and another concave: its name thus derives from a resemblance to the new moon. In the fragment, Hippocrates was able to calculate the area of various lunes, which at the time meant to quadrate them: that is, to construct a square whose area was of equal magnitude.

Consider the diagram below. From Pythagoras' Theorem and the fact that angles in a semicircle are right angles, we have that $AC^2 = AB^2 + BC^2 = 2AB^2$. Since Hippocrates was willing to assume the principle that the areas of circles are in the same ratio as the areas of squares constructed upon their on their diameters – which was probably only proved later by Eudoxus, using what later became

[65] Marcus Giaquinto, "Visualizing in Mathematics", in *Philosophy of Mathematical Practice*, ed. Paolo Manders (Oxford: Oxford University Press, 2011), 22-42.
[66] Though it may have received some modification from Simplicius, who preserved the fragment from Eudemus' *History*. See Jacques Brunschwig and Geoffrey Lloyd, ed., *The Greek Pursuit of Knowledge* (Harvard: Harvard University Press, 2003), 245.

known as the 'method of exhaustion' – it follows that the semicircle with diameter AC has twice the area of the semicircle with diameter AB. But the quarter-circle $AOBF$ has half the area of the former semicircle, and so its area is equal to that of the latter semicircle. As the segment ADBF is common to both, it follows that the two shaded regions are of equal area: that is, the area of the lune $AFBE$ is equal to that of the triangle AOB. This is in turn equal to the area of a square constructed upon the line segment AD (not drawn).

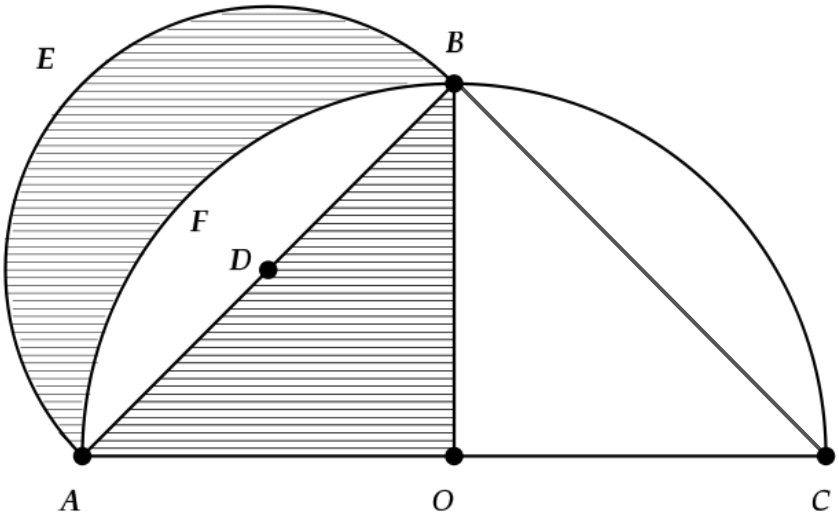

Diagram 2.1. The figure AFBE is amongst the most famous of Hippocrates' Lunes. Its area is equal to that of a rectilinear figure: the shaded triangle AOB.
The diagram is taken from Piers Bursill-Hall's unpublished lecture notes from a History of Mathematics course at the University of Cambridge.

Amongst those figures whose area Hippocrates was able to find is the one given in Diagram 2.2 below: both the thin lune at the top and the shaded circle at the centre were quadrated at the same time.

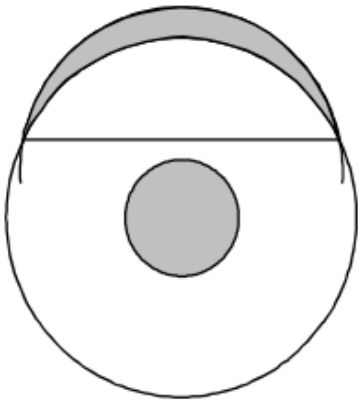

Diagram 2.2. Hippocrates used a very complicated procedure to square this lune and circle together.
Piers Bursill-Hall

An important mathematical problem at the time was to quadrate or 'square' the circle. As said above, the Ancient Greek conception of area was essentially tied to rectilinear area, so the fact that no one had been able to solve this problem was a cause for concern. Its insolubility might have indicated that their whole concept of area was problematic when applied to curvilinear shapes.

Aristotle later accused Hippocrates of thinking he had squared the circle, having shown how to square some individual lunes as well as a lune and a circle together.[67] Of course, this would have been a mistake, though it seems implausible that a mathematician of Hippocrates' stature would make such a careless error, and Aristotle's accusation is perhaps unfair. Yet the quadration of some lunes on their own, as well as other lunes together with a circle, would have seemed to suggest that areas of circles are in some sense analogous to areas of lunes, furnishing a kind of heuristic, philosophical argument by analogy that the Greek conception of area can be extended to circles. However, it turns out that only some lunes can be squared by ruler-compass construction, whereas others can't. In 1882, the Lindemann-Weierstrass Theorem put an end to any hopes of squaring the circle using ruler-compass methods by showing that π is transcendental, so quadrature of the lune in Diagram 2.2 by itself must indeed be impossible.

Non-deductive methods such as given in these examples have always been of great use to mathematicians, both in contemporary research and throughout history. Practitioners often arrive at conjectures using heuristic techniques; without a conjecture to aim at it is again usually far from clear how to go about employing deductive methods. Moreover, we shall see in the next section that the scope of non-deductive methods often extends to discovering proofs as well. Consider also a few quotations by some eminent mathematicians:

> 'Even in the mathematical sciences, our principle instruments to discover the truth are induction and analogy.' – Laplace[68]
>
> 'In the Theory of Numbers, it happens rather frequently that by some unexpected luck, the most elegant new truths spring up by induction.' – Gauss.[69]
>
> 'As we must refer the numbers to the pure intellect alone, we can hardly understand how observations and quasi-experiments can be of use in investigating the nature of the numbers. Yet in fact, as I shall show here with very good reasons, the properties of the numbers known today have been mostly discovered by observation, and discovered long before their truth has been confirmed by rigid demonstrations.' – Euler[70]

Focusing on finished pieces of mathematics may lead us to overlook this importance, however: only the deductive proof is sent for publication, and the heuristic work is then discarded.

---

[67] Aristotle, "Sophistical Refutations: Book 11", in *The Complete Works of Aristotle, Volume 1*, trans. Jonathan Barnes (Princeton: Princeton University Press, 1984), 291.
[68] Quoted in Pólya, *Mathematics and Plausible Reasoning, Volume 1*, 35.
[69] Ibid., 59.
[70] Ibid., 3.

As well as their role in discovery, a second use of non-deductive methods is as a check on our deductive work. In Section 1.v we noted that this is particularly important when the foundations or underlying concepts have not yet been clarified, and consequently some key techniques have not yet been shown to be sound. Here we mentioned Euler's work on infinite series and the physical applications of the calculus. Yet even when the theoretical foundations have been secured, non-deductive methods can provide useful corroboration: 'Generally, our mathematical certainty does not rest exclusively on either logico-deductive methods or quasi-empiricism but on a healthy combination of both.'[71] This is especially important for the kinds of cases we pointed out in Section 1.vi, where it is difficult to tell if a very long or computer-based argument is sound.

2.ii Non-deductive Techniques in the Context of Discovery

Let us focus the discussion by concentrating on a particular problem.

Problem 2.1.

> Alison and Josh are standing at the top of a flight of 10 stairs. Alison can jump down one or two stairs at a time (for example, she could get to the bottom by taking five jumps of two stairs at a time). In how many distinct ways can Alison get from the top of the stairs to the bottom? Josh can jump up to 10 stairs in one go. In how many distinct ways can Josh get from the top of the stairs to the bottom?[72]

We interpret the question in the natural way, to mean there is a landing at the top and the bottom of the stairs, both of which they must each visit, and 9 distinct flat standing areas on the intermediate steps, none of which are essential to visit, so that there are 10 vertical sections in total. There does not seem to be anything special about the number 10, so let us generalise the problem to n steps and then focus on simpler cases. Let $A(n)$ be the number of ways in which Alison can climb down a flight of n stairs (i.e., a flight of stairs with $(n-1)$ distinct platforms in addition to the upper and lower landings).

If we are not sure how to proceed in calculating $A(n)$, we can begin by generating some numerical data. We note by enumerating all cases that $A(1) = 1, A(2) = 2$ and $A(3) = 3$. So far this is consistent with a number of conjectures, such as $A(n) = n$ for all n, but if we calculate the fourth term $A(4) = 5$ this solution is excluded. When we work out $A(5) = 8$, some readers might spot a pattern emerging: the Fibonacci numbers! At this stage we could check by calculation that $A(6) = 13$, but actually I already feel fairly convinced of the conjecture that $A(n) = F_{n+1}$ for all n, and therefore that it will be a better use of time to now set about proving the conjecture, rather than generating any more data. The Fibonacci numbers are characterised by the recurrence relation $F_n = F_{n-1} + F_{n-2}$ and so as the initial terms match up it suffices to prove that

[71] De Villiers, "The Role and Function of Quasi-Empirical Methods in Mathematics", 411
[72] UKMT Senior Mentoring Scheme, 2013-2014, Sheet 1, Q3.

$$A(n) = A(n-1) + A(n-2) \text{ for } n \geq 3.$$

Aha! If Alison begins her journey by traversing just 1 step, she has $A(n-1)$ unique ways to finish her trip, whereas if she starts with 2 steps, there are $A(n-2)$ unique ways to finish off. No combination of steps could feature in both categories, beginning with both 1 and 2 steps, and for her a journey of any length must begin with a single or double step, so the result we need follows. Lastly, we calculate $A(10) = 89$.

This solution is a typical example of the quasi-empirical approach: looking at simpler cases, getting stuck in and working out a few examples, trying to find a pattern that can be exploited, but ultimately not knowing in advance where this search will lead. This can be contrasted with a more considered, self-aware mode of proceeding where a definite strategy is always present. An experienced problem solver would probably begin by using recursion straight away, and might regard the somewhat cumbersome and messy inspection of more individual cases than strictly necessary as a last resort to be used only when other options had been exhausted. Let us now consider the case of Josh.

We may again proceed quasi-empirically. We work out a few values of $J(n)$, the number of ways Josh can traverse the n stairs if he is able to jump up to n stairs at a time, in order to acquire some data as the basis for a conjecture. We note that again $J(1) = 1$ and $J(2) = 2$, but this time have that $J(3) = 4$ and $J(4) = 8$. This immediately suggests the conjecture $J(n) = 2^{n-1}$. We could check by enumerating cases that $J(5) = 16$, but at this stage I again have a strong feeling that this is a waste of time and effort and that it is instead more profitable to go straight to attempting to prove the conjecture. The degree of felt confidence is remarkable given that we have only checked a few small values!

We may recall that that 2^{n-1} is the number of subsets of a set of size $(n-1)$, or the number of ways to make $(n-1)$ independent binary choices. There are also $(n-1)$ intermediate platforms, each of which Josh can choose to land on or not. Aha! The number of ways of climbing down the stairs is just the number of ways Josh can select which of the $(n-1)$ optional step-platforms he would like to land on, whilst avoiding the rest. Having spotted this proof of our conjecture, we can now calculate that $J(10) = 512$.

An alternative route through the second part of Problem 2.1 would be to look for a recursion straight away, by analogy with the first part of the problem. Similarly to before, we have that $J(n) = J(n-1) + J(n-2) + .. + J(1) + 1$ for $n \geq 2$, depending on whether Josh makes an initial leap of 1,2,3, ... , or $(n-1)$ steps, or instead jumps down the whole flight of steps in one go, which can be done in exactly 1 way.

Again, in order to proceed from this recursion to a closed expression for $J(n)$ we might work out a few values and then notice that it doubles each time. Alternatively, it might occur to us that because $J(n-1) = J(n-2) + .. + J(1) + 1$ we can write $J(n) = J(n-1) + [J(n-2) + .. + J(1) + 1] = 2J(n-1) =$

$2^{n-1}J(1) = 2^{n-1}$ as $J(1) = 1$. This way of proceeding is perhaps more pleasing than making use of quasi-empirical data, although in this case it involved a bit more work than the previous argument. The final proof based on subset selection was very concise indeed.

Let us now make a few observations that connect this example to the central themes of this chapter. Firstly, one interesting feature of these solutions is that the quasi-empirical work not only led us to discover the answer but helped us find the proof as well. When the quasi-empirical work yielded the first few terms of the Fibonacci sequence in the first part, this suggested using recursion. When it led us to the formula 2^{n-1} in the second part, our earlier observation that there were $(n-1)$ distinct platforms between the two landings gave an opportunity to spot the crucial analogy with constructing subsets. This is reminiscent of a famous quote by Riemann, suggesting that getting to the answer is often the most difficult part: 'If only I had the theorems – then I should find the proofs easily enough!'[73]

Another advantage of carrying out the quasi-empirical work was that it provided an important check on our deductive counting arguments. Teaching experience discloses that students who are new to learning combinatorics usually make a large number of mistakes that must be corrected with training and practice. These include counting members of a set multiple times when enumerating the members of that set, or giving a formula that breaks down for boundary cases such as $n = 1$. For such students, although perhaps only to a lesser extent for more experienced problem-solvers, the inductive work of checking a few special cases directly will increase their confidence in the result. And surely this is a reasonable procedure for them to employ, and the corresponding increase in confidence is rational. Nevertheless, it is perhaps the mark of a very experienced problem solver not to bother with any such calculations. So it would seem that how much quasi-empirical work it is reasonable to do as a check depends on the mathematical experience and ability of the enquirer.

We can also see that in both cases the reasonableness of the decision to stop calculating more data and turn instead to attempting to prove a conjecture hinged on having gained at least some degree of confidence in the truth of that conjecture, based on the hope that the observed pattern would continue. Reaching the correct decision here can be very important: both the process of generating more data unnecessarily and attempting to prove a false conjecture will result in wasted time and energy. This is even more important when considering the much more complicated problems professional mathematics study, and here perhaps a research grant or even a career might be at stake. So the question of whether a body of quasi-empirical evidence warrants investing time looking for a proof is of central importance for day-to-day mathematical research.

Yet although a decision is required here, the choices we make will seldom if ever be unequivocal. If the reasonableness of switching to looking for a proof after producing my initial data had been challenged, or if someone did not share my intuitions about the prior plausibility of the conjecture for the first part, I would be

[73] Quoted in Imre Lakatos, *Proofs and Refutations* (Cambridge: Cambridge University Press, 1976), 9.

at a loss as to how to convince them. After all, only five terms were calculated here, and we had already dismissed a failed conjecture that worked for the first three. There is simply no way in which the reasonableness of our confidence at this stage could have been established with anything like the finality with which the proof later supplied established the truth of the conjecture itself. Of course, in practice I may never be challenged: usually I will simply make a private judgement about how to proceed. This suggests that this kind of inference is more suited to the context of Private Acceptance rather than Public Acceptance. Then if I am wrong only my own time and energy is wasted. If I am correct, only the final proof will be given, and the heuristic work will go unpublished.

2.iii. Computers and Non-Deductive Methods as Warrant

In Section 1.vi of the last chapter, we saw how the use of computer-based methods has impacted on the effectiveness of proof. In contrast, I will in this section show how the potential for non-deductive methods has been greatly enhanced over the last few decades by the increase in available computing power they have supplied. Indeed, one may even go so far as to liken the influence of the computer upon quasi-empirical investigation in mathematics to the telescope affecting cosmology or the microscope biology and chemistry. For the two examples given in this section, the non-deductive work will be sufficiently compelling to make us strongly inclined to believe the results they lead to. This moves us beyond mere discovery and into the context of justification.

Compared to the unaided human intellect, the computational power supplied by modern computers is in many respects vast. In number theory, mathematicians can verify conjectures for huge numbers of integers in a fraction of a second with only the press of a few buttons, and easily perform calculations that often would have been intractable or at least taken inordinate amounts of time a few decades ago. Programs such as the PSLQ integer relation algorithm mentioned earlier have been developed to spot numerical patterns automatically. Websites such as Wolfram Alpha supply users with a wide range of information relating to mathematical input of any description, be it numerical or qualitative.[74] And analogous graphical techniques have also been developed, although here the human visual system is still an impressive instrument.

One especially interesting development in computer methods is the geometrical environments alluded to in Section 2.i, which include *Cabri*, *Sketchpad* and *Cinderella*. These programs enable geometers to set up a single interactive diagram on a computer and freely alter variables such as the direction or length of a line segment, the radius of a circle, or the position of a point. In doing so, the possible range of variations of a diagram falling under a given specification can be explored far more efficiently than by relying on traditional pencil-and-paper methods. The diagrams these programs produce are very accurate and can be constructed quickly and easily. A more recent application is *Apollonius*, which we will now use to tackle another concrete problem.

[74] http://www.wolframalpha.com

Problem 2.2.

Circle Υ_1 lies inside circle Υ_2 and touches it at A. From a point P (distinct from A) on Υ_2, chords PQ and PR of Υ_2 are drawn touching Υ_1 at X and Y respectively. Show that $\angle QAR = 2 \times \angle XAY$.[75]

Firstly I used *Apollonius* to construct an interactive diagram, following the set-up described in the question (labels added separately).

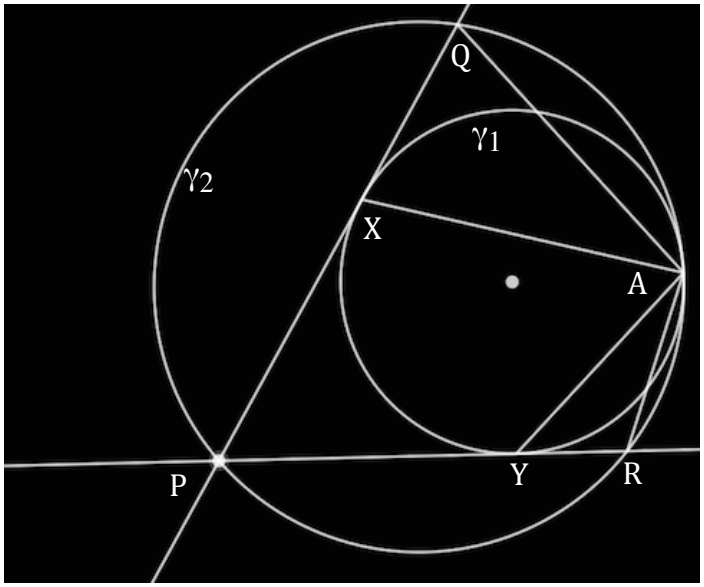

Figure 2.3. Implementation of problem 2.2 in Apollonius

The point P and the unlabeled white dot marking the centre of Υ_1 can be moved interactively to alter the diagram. After playing around with these parameters for a while, it became visually apparent that a stronger claim was also true: joining P and A, I guessed that $\angle PAY = \angle YAR$ and $\angle PAX = \angle XAQ$. This claim is stronger in the sense that the theorem to be proved follows immediately from it, but also simpler in that it comprises two smaller claims we can prove separately and which involve only triangles rather than quadrilaterals. Hence, the proof of them is likely to be fairly easy to find, given that the stronger claim is true. After adding the line segment AP on Apollonius, varying the position A and the radius of Υ_1 results in a visual experience that very strongly inclines one to believe that the new conjecture is indeed true. Unfortunately I cannot reproduce this experience for the reader here, but have included a few static images instead:

[75] UKMT Senior Mentoring Scheme, 2012-2013, Sheet 1, Question 6

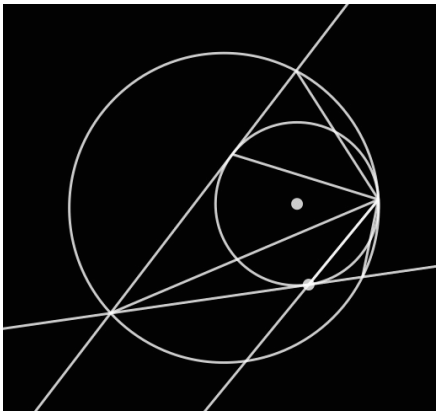

Figure 2.4

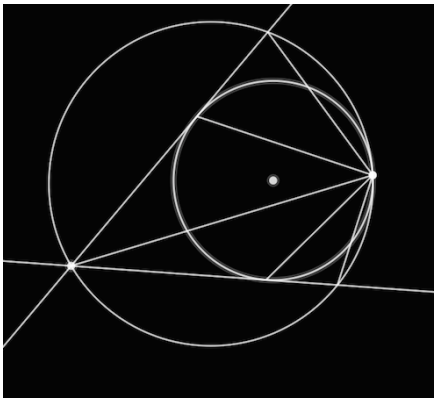

Figure 2.5

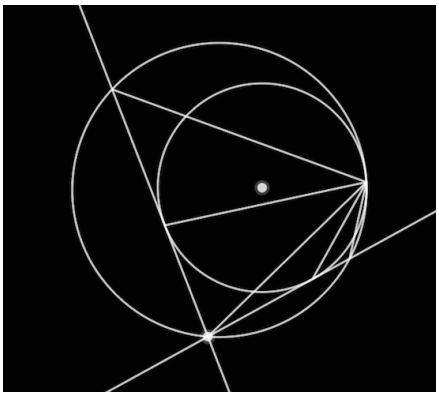

Figure 2.6

Let us now attempt to prove the stronger claim using circle geometry. Let Z be the point where AP meets Υ_1, as in the hand-drawn diagram below, where I have also added the line YZ to enable us to make use of the Alternate Segment Theorem later on. Lastly, we draw in the mutual tangent to both circles at A, and define T as some arbitrary point on the lower half of this tangent (i.e. on the opposite side to Q of the line segment PA produced, as shown). Let $\angle PAY = x$ and $\angle APR = y$.

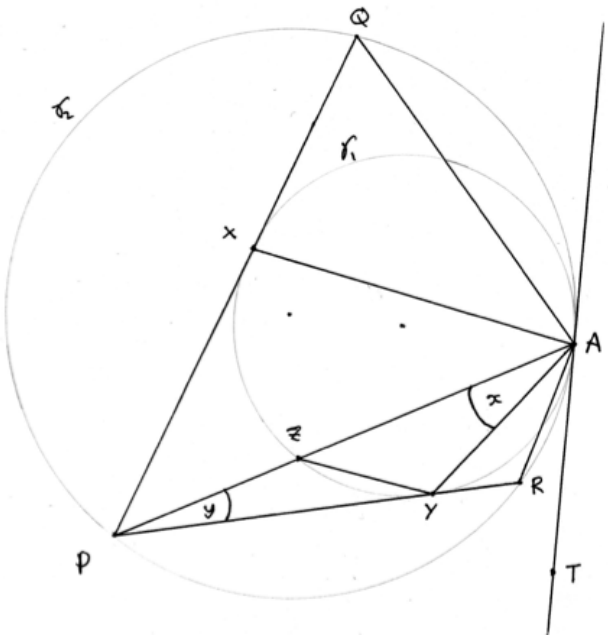

Figure 2.7. Scanned drawing of Problem 2.2 with constructions.

By comparing the two triangles ARY and APR we have that the angle $\angle AYR$ is equal to $\angle APY + \angle PAY = x + y$, because both must yield 180° when added to the sum of $\angle ARY$ and $\angle RAY$. Using the Alternate Segment Theorem on Υ_1 with the tangent line PR and the triangle AYZ gives us that the angle $\angle AZY$ is equal to $\angle AYR$, and so is also equal to $x + y$ from before. This is also the size of $\angle YAT$ by the same theorem applied with Υ_1, the tangent line AT, and the same triangle AYZ. Lastly, applying the Alternate Segment Theorem with Υ_2, the tangent line AT, and the triangle APR tells us that the angle $\angle RAT$ is equal to $\angle APR = y$, so that angle $\angle YAR = \angle YAT - \angle RAT = (x + y) - y = x = \angle PAY$ as required. The proof that $\angle QAX = \angle PAX$ is almost identical.[76]

Let us now consider another example, this time from number theory.

Problem 2.3.[77]

Does there exist a positive integer, n, such that $(2 + \sqrt{2})^n$ differs from an integer by no more than 10^{-6}?

My initial thoughts upon reading this are that such integers will exist because the distance from each element of the sequence to the nearest integer is likely to take random values from the interval $(0,1/2)$ as we multiply by a non-integer at each step. However, it is not clear how to start looking for a rigorous argument for this

[76] Enthusiastic readers are recommended to obtain a version of Apollonius and use this approach to solve problems 2 and 6 on sheet 3 of the senior mentoring scheme, 2014-2015.
[77] UKMT Senior Mentoring scheme, 2013-2014, Sheet 1, problem 8.

claim. So let us instead begin by using a *Microsoft Excel* spreadsheet to calculate the first few values of the expression and then looking for a pattern.

n	$(2+\sqrt{2})^n$
1	3.4142135624
2	11.6568542495
3	39.7989898732
4	135.8822509939
5	463.9310242292
6	1583.9595949289
7	5407.9763312574
8	18463.9861351716
9	63039.9918781715
10	215231.9952423430
11	734847.9972130290
12	2508927.9983674300
13	8566015.9990436600
14	29246207.9994398000
15	99852799.9996718000

Examining these data, we see that from the 5th row onwards the decimal expansion always begins with .9, from the 9th row .99, and is then .999 in the final three rows. So looks as though my initial guess was correct, but for the wrong reasons: it appears that $s(n) = (2+\sqrt{2})^n$ actually gets progressively closer to an integer at every step! In another case, this kind of heuristic argument might have led us to the wrong answer here. After making an initial attempt to find a formula in terms of n for the integer nearest to the value of the expression, which was unsuccessful, it occurred to me to calculate the difference between the expression and the next integer up – as for the calculated values in the table the expression is always less than the integer closest to it (apart from the first term). Call this difference $f(n)$.

n	$(2+\sqrt{2})^n$	$f(n)$
1	3.4142135624	0.5857864376
2	11.6568542495	0.3431457505
3	39.7989898732	0.2010101268
4	135.8822509939	0.1177490061
5	463.9310242292	0.0689757708
6	1583.9595949289	0.0404050711
7	5407.9763312574	0.0236687426
8	18463.9861351716	0.0138648284
9	63039.9918781715	0.0081218285

10	215231.9952423430	0.0047576570
11	734847.9972130290	0.0027869712
12	2508927.9983674300	0.0016325708
13	8566015.9990436600	0.0009563416
14	29246207.9994398000	0.0005602203
15	99852799.9996718000	0.0003282130

Looking at these figures we might also notice that they seem to decrease by around the same proportion each time, suggesting that we should calculate the ratio between terms:

n	$(2+\sqrt{2})^n$	$f(n)$	$f(n)/f(n-1)$
1	3.4142135624	0.5857864376	
2	11.6568542495	0.3431457505	0.5857864376
3	39.7989898732	0.2010101268	0.5857864376
4	135.8822509939	0.1177490061	0.5857864376
5	463.9310242292	0.0689757708	0.5857864376
6	1583.9595949289	0.0404050711	0.5857864376
7	5407.9763312574	0.0236687426	0.5857864376
8	18463.9861351716	0.0138648284	0.5857864378
9	63039.9918781715	0.0081218285	0.5857864387
10	215231.9952423430	0.0047576570	0.5857864458
11	734847.9972130290	0.0027869712	0.5857864789
12	2508927.9983674300	0.0016325708	0.5857867410
13	8566015.9990436600	0.0009563416	0.5857887761
14	29246207.9994398000	0.0005602203	0.5857951978
15	99852799.9996718000	0.0003282130	0.5858640937

Initially, the ratio is always equal to the first term $0.5857864376\ldots$, although this pattern changes somewhat later in the list. Examining the rest of the table, it seems that we can attribute this to a rounding error of some sort. The zeros that appear at the end of the numbers from the 10^{th} line onwards suggest that only about 15 decimal digits are stored in total.

At this stage, I spotted that 0.5857864376 is the beginning of the decimal expansion of $(2-\sqrt{2})$. If $(2-\sqrt{2})$ is in fact the exact value of both the first term and the ratio of each pair of adjacent terms, then $f(n)$ must be equal to $(2-\sqrt{2})^n$ for each positive integer n. We therefore arrive at the following conjecture:

$(2+\sqrt{2})^n + (2-\sqrt{2})^n$ is an integer for all positive integers n.

This claim can now proved by considering the binomial expansion of the two bracketed expressions. Terms with an even power of $\sqrt{2}$ will themselves be integers, and terms with an odd power of $\sqrt{2}$ in the first expression will be cancelled out by corresponding terms in the second expression, which will be identical except for having a negative sign. So all we need now do to solve Problem 2.3 is to pick n such that $(2-\sqrt{2})^n < 10^{-6}$. This can easily be done with logarithms ($n = 26$ is the first solution).

Let us now reflect on these two examples and see what conclusions can be drawn with regards to the central questions of this chapter. Firstly, consider the geometry problem solved using *Apollonius*. By varying the parameters continuously I was able to spot an invariant, and this led to a related conjecture that turned out to be easy to prove. Of course, without experience in solving geometry problems I may not have known to capitalise on this discovery. But without Apollonius I would not have noticed the invariant at all: it stood out visually because everything else was changing around it as I varied the parameters.

When combined with the developed intuitive and visual abilities of a specialist geometer, this new mode of discovery – varying the parameters and looking for invariants or other interesting patterns – can often lead to results that would otherwise be very hard to come by. For example, Adrian Oldknow has used the application *Sketchpad* to discover a number of results including that the Soddy centre, incentre and Gergonne point of a triangle are collinear, and June Lester has used the same application to discover that the two Fermat points, the nine-point centre, and the circumcentre of a triangle always lie on a circle (now know as the Lester circle). De Villiers has also discovered generalizations of the Fermat-Torricelli point of a triangle and of Neuberg's Theorem.[78] As a consequence of these kinds of applications, in 2004 Villiers stated that research in traditional Euclidean geometry was undergoing an 'exciting revival', and mentioned Philip Davis' prediction of a 'resurgence' in triangle geometry.[79]

Experience of teaching students to construct geometric proofs, such as the proof required in Problem 2.2, discloses that a common mistake is to rely on features that do not belong to the problem essentially, but rather are only specific properties of the diagram that the student happens to have drawn to represent it. Programs like *Apollonius* can therefore rationally increase our confidence in our geometrical reasoning, because we can continuously deform the diagram until we have had a visual experience that has included representatives of all the possible kinds of cases that might have invalidated it.

So far this fits with the traditional roles of non-deductive methods in mathematical research outlined earlier: a non-deductive check on our deductive methods. However, even prior to the discovery of a proof the pull of the visual experience towards believing the conjecture was already quite strong, especially after I had drawn in the construction line AP. And it feels as though this time the induced confidence is reasonable. After all, I varied each of the parameters across the

[78] De Villiers, "The Role and Function of Quasi-Empirical Methods in Mathematics", 401.
[79] Ibid, 401.

entire possible range of their values, thus checking a large number of diverse representative cases, and the stated invariant appeared to hold true throughout.

In a paper that specifically addresses this issue, the geometer Michael Fox describes what he thought at first was a generalisation of the Six-Circles Theorem for pentagons with an incircle, which he announced at the Mathematical Association's Annual Conference in 2006. Sadly, the conjecture – which would have been a beautiful theorem – turned out to be false. John Silvester later pointed out that although the two circles Fox conjectured to be touching always visually appeared to do so as he varied the parameters, they are actually offset by a tiny amount, below the threshold of visual perception (the discrepancy is reported as about 0.12%). Fox writes, 'The lesson should be clear: *in geometry do not believe your eyes*.'[80] Hence, the visual argument for Problem 2.2 does not rationally compel assent with the same authority as the proof that is later supplied.

With access to *Microsoft Excel* available, finding the solution to Problem 2.3 was rather easy and indeed almost mechanical. The only interesting parts were deciding to look at the ratio of successive terms and spotting the decimal expansion of $2-\sqrt{2}$. The process was greatly facilitated by *Excel*'s built-in devices like relative cell references: just by clicking and dragging a formula we can create a whole column of data. Without access to a computer, it is likely that I would never have bothered with this particular quasi-empirical approach, even if I had the latest scientific calculator. I would have had to think of the proof using the Binomial Theorem from scratch, and this might have taken me much longer.

In some ways Problem 2.3 is again a fairly typical example of the received view of mathematical practice. The non-deductive work both provided us with the right answer and led us to the discovery of the proof itself, though it will not feature in the final presentation of the argument, which may simply begin from the observation that $(2+\sqrt{2})^n + (2-\sqrt{2})^n$ is an integer for all positive integers n. However, as before the quasi-empirical work also moved us beyond merely discovering a conjecture that was sufficiently plausible to warrant further investigation. The fact that $f(n)$ decreases by a factor of roughly $(2-\sqrt{2})$ for all of the values checked seems like compelling evidence that the conjecture was true, and even more so given that the similar expression $(2+\sqrt{2})$ occurs prominently in the question.

Nevertheless, until we reach the point where a specific integer with the requisite property has actually been constructed, the confidence that can be rationally induced by the data still falls short of complete certainty. It is possible – although perhaps highly unlikely – that such patterns arise simply by chance. This tendency to expect a pattern to continue can be sometimes be exploited, however: it is famously easy to fool people with little experience of number theory into thinking that the expression $f(n)=n^2+n+41$ is prime for each positive n, even though it is clearly composite whenever $41|n$, because surprisingly it is in fact prime for $n=1,2,..,39$, and they are unlikely to check values outside of this range.

[80] Michael Fox, "Using the Geometer's Sketchpad, Part 1: General Issues", *Mathematical in School*, 37 (2008), 3.

In both these cases we have moved beyond simply making an intelligent assumption for the practical purpose of deciding what to do next. For both problems in this section I felt it appropriate to describe my own psychological state, which was induced by the non-deductive evidence alone, using the term 'belief', intended in its fullest sense. Yet as psychologically convincing as this evidence was, in both cases it lacked the finality that came with explicit proof. Although we might expect the effect on mathematically literate adults to be one of belief, if someone were to dissent from our judgement we could not convict them of straightforward irrationality as we could if they were to knowingly reject a deductively valid argument, even if we might be surprised at their reaction. This given, the non-deductive evidence will be less effective in securing Consensus amongst the readership of such an argument.

2.iv. The Influence of Background Knowledge

Let us investigate another example. This time we will see that the apparent evidential import of the same body of non-deductive evidence for a claim seems to vary with the degree of background knowledge of the enquirer. Again, this shows that although some individuals may find a body of non-deductive evidence compelling, other with a different level of experience or background knowledge may be less convinced. Non-deductive techniques of this nature are therefore less effective in securing Consensus amongst mathematicians.

Problem 2.4.

Consider the set of regular polygons in the plane, with fixed perimeter p. Which of these has the largest area?

This is sometimes called the 'fencing problem' because it might be motivated by considering the task of determining the largest area a farmer can enclose in a field using only a certain amount of fencing.

We begin by calculating the area of enclosed by a regular polygon with n sides. By drawing construction lines connecting each vertex to the centre of the enclosed region, we can divide it into n isosceles triangles with the unequal side (unless $n = 6$, whence the triangles are equilateral) equal to p/n and subtended by an angle of $2\pi/n$ radians. We can again split each of these into two right-angled triangles by joining the midpoint of the unequal side to the opposite vertex. This gives us $2n$ right-angled triangles with dimensions as follows:

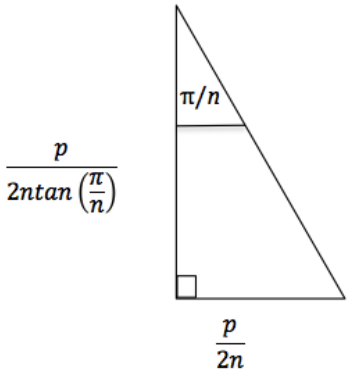

Figure 2.8. The polygons are each composed of $2n$ triangles congruent to (or reflections of) the one given here.

The area of this triangle is $\frac{p^2}{8n^2\tan\left(\frac{\pi}{n}\right)}$ and of the polygon therefore $\frac{p^2}{4n\tan\left(\frac{\pi}{n}\right)}$.

As p^2 is a constant that can be factored out of the expression, we will from now on assume without loss of generality that it is equal to 1 (this simplification could also have been made immediately by scaling the diagram). As before, we can begin quasi-empirically, by calculating the area of regular n-gons for a few different values of n.

n	**Area**	***n***	**Area**
3	0.048112522	13	0.078022298
4	0.062500000	14	0.078237255
5	0.068819096	15	0.078410502
6	0.072168784	16	0.078552180
7	0.074161478	17	0.078669522
8	0.075444174	18	0.078767803
9	0.076318817	19	0.078850940
10	0.076942088	20	0.078921894
11	0.077401983	21	0.078982935
12	0.077751058	22	0.079035827

It seems from the data that the area is strictly increasing with n. If this pattern continues then Problem 2.4 must be a trick question. There will be no single answer, but only an infinite sequence of polygons whose area strictly increases:

Conjecture 2.5.

For $n \geq 3$, let p_n denote the area of a regular polygon with n sides and of unit perimeter. Then $n > m \implies p_n > p_m$

We have only looked at a small number of cases, so giving our assent to this conjecture might be premature. Let us test it for a few larger values.

n	**Area**	***n***	**Area**
1000	0.079577210	2000	0.079577406
1100	0.079577255	2100	0.079577412
1200	0.079577290	2200	0.079577417
1300	0.079577317	2300	0.079577422
1400	0.079577338	2400	0.079577426
1500	0.079577355	2500	0.079577430
1600	0.079577369	2600	0.079577433
1700	0.079577381	2700	0.079577436
1800	0.079577391	2800	0.079577438
1900	0.079577399	2900	0.079577440

The pattern has indeed continued for the checked values. After producing these data I then used an *Excel* spreadsheet to check the pattern for every value of n from 3 up to 10,000. The pattern continued throughout this entire range, strongly suggesting that the sequence p_n is indeed given by an increasing function of n.

Perhaps one reason why the numerical evidence is particularly convincing here is because it seems plausible that – to put the matter crudely – all that 'matters' is the 'size' of n, its brute numerical magnitude. However, more number-theoretic properties of n are relevant to some geometrical questions about polygons. In his *Disquisitiones Arithmeticae*, published in 1801, Gauss showed that an n-sided polygon is constructible using only ruler and compass if n is the product of a power of 2 and any number of distinct of Fermat primes, and correctly conjectured that this was also a necessary condition.[81]

In response to this reservation, someone who was still firmly convinced by the heuristic argument might say: 'Such number-theoretic properties could be mathematically relevant in some sense, yes, but not to its *area*.' Again, it would not be *that* surprising to find systematic correlations between number-theoretic properties of n and properties of p_n: whether the latter is a quadratic irrational, say. 'Yes, but not as to the *magnitude* of the area!' they may retort. But it is not clear there are sufficient grounds for this claim; it is always hard to know in advance which properties will be relevant. We shall see in the next section that many patterns that continue for a very long time nevertheless break down eventually. At this stage there could still be other possibilities we have not yet thought of; a reader might even be better informed than we are here and so for these reasons find the argument unconvincing.

Let us now consider how judgements about the import of the evidence are likely to be affected by having prior knowledge of two related results. The first such result we shall consider is the Isoperimetric Theorem, mentioned in Section 2.i:

Theorem 2.6. The Isoperimetric Theorem

The area of a plane figure having fixed perimeter p is maximal if and only if the figure is a circle.

Again, we may restrict our attention to the case $p = 1$. Now, we can supplement the numerical findings with an observation easily supplied by imagination. As n gets larger, the polygon will eventually become visually indistinguishable from – but never actually identical to – a circle with circumference 1. Moreover, the 'degree of resemblance' also increases with n. A circle of perimeter 1 has an area of $1/(4\pi) = 0.0795774715459478\ldots$ which is in fact very close to – although of course slightly larger than – the last value calculated in our table of polygon

[81] A Fermat prime is a prime of the form $2^{2^n} + 1$. A proof of necessity was published in 1837. Pierre Wantzel, "Recherches sur les moyens de reconnaître si un Problème de Géométrie peut se résoudre avec la règle et le compas." *Journal de Mathématiques Pures et Appliquées* 1 (1937): 366–372.

areas. The Isoperimetric Theorem tells us that $p_n < 1/4\pi$ for each n, and putting this together with the numerical and visual data the conjecture that $\langle p_n \rangle$ is a strictly increasing sequence, perhaps now tending to $\pi/4$, becomes even more seductive. Moreover, had we known this in advance (and again perhaps the reader is in this position already), it is reasonable to think that we would have found the quasi-empirical data more convincing to begin with.

We now consider the impact of knowing a second result. To derive this, we use a procedure employed by Eudoxus in the first half of the 4th century BCE, following from the ideas of Antiphon and Bryson, and which is now known as the 'method of exhaustion' (although this name dates from the seventeenth century, and it is unlikely that the Greeks had anything so explicit and coherent to warrant the label 'method').[82] Eudoxus was able to use this technique to prove a number of results in plane and solid geometry, including the theorem that the areas of circles as in the same ratio are as squares constructed on their diameters, which was mentioned in connection with Hippocrates in Section 2.i. Though we will use the same construction, we proceed in a rather different direction, using modern techniques. Consider the following proposition, a reformulation of Euclid X.1.

> 'If from any magnitude there be subtracted a part not less than its half, and if from the remainder one again subtracts not less than its half, and if this process of subtraction is continued, ultimately there will remain a magnitude less than any preassigned magnitude of the same kind.'[83]

Now, we begin with an equilateral triangle circumscribed by a circle of unit circumference. Let the area of this be c_1. We construct a sequence $\langle c_n \rangle$ inductively by doubling the number of sides of the polygon inside the circle, as in the diagram below, and taking their successive areas.

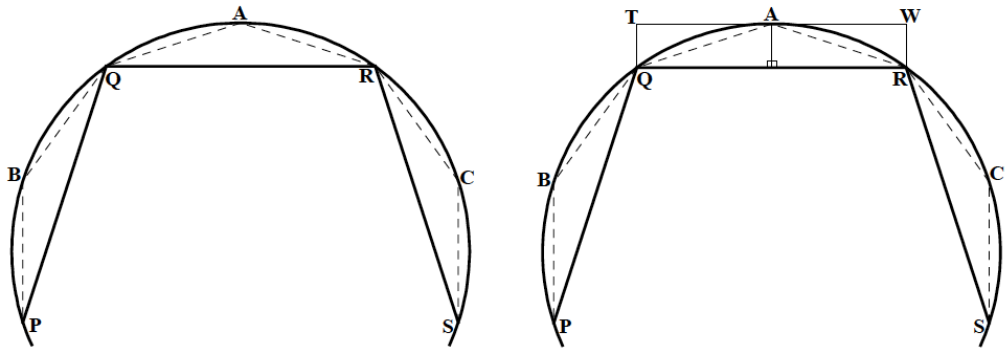

Figure 2.8. Each of the sides is replaced with two equal, shorter sides.
Here QR is replaced with QA and AR, each touching the circle at A.
Piers Bursill-Hall.

[82] Pascal Boyer, *The History of the Calculus and its Conceptual Development* (New York: Dover Publications, 1959): 33-34, 36.
[83] Uta Merzbach and Carl Boyer, *A History of Mathematics* (New Jersey: John Wiley and Sons), 82

We can see from the diagram that the difference between the areas of the polygons and that of the circle decreases by a factor of more than two each time. Although the Greeks hadn't our conception of a limit, we can easily see from Euclid X.1 that

$$\lim_{n\to\infty} c_n = \pi/4$$

Because the shortest distance between two points is a straight line, it follows that the perimeter of each polygon is less than 1. So each of these polygons is smaller in area to the corresponding enlarged polygon with perimeter 1, and it follows that $c_n < p_{3\times 2^{n-1}}$ for each n. By the sandwich rule for the convergence of sequences and the Isoperimetric Theorem, we also have that

$$\lim_{n\to\infty} p_{3\times 2^{n-1}} = \pi/4$$

Moreover, the limit approaches from below, so there must be a subsequence that is strictly increasing. This gives us the following result:

Theorem 2.7.

There exists a subsequence $\langle p_{j(n)} \rangle$ such that $n > m \Rightarrow p_{j(n)} > p_{j(m)}$.

Though Conjecture 2.5 is not yet proved, this theorem – which establishes that some infinite subsequence of the regular polygons has the property expressed in the conjecture – makes the numerical data given earlier even more compelling. And again, it is likely that someone who already knew Theorem 2.7 would have found the initial evidence for Conjecture 2.5 more convincing, all other things being equal, than a second enquirer who did not.

Let us now draw some conclusions from the results of this section. When Theorems 2.6 and 2.7 were taken into account, the non-deductive data became progressively more convincing. Hence it is likely that individuals who had prior knowledge of or perhaps even substantial experience working with these or other related claims would in consequence have automatically been more convinced by the quasi-empirical data, because Conjecture 2.5 would have seemed more plausible in and of itself (the reader may indeed have found themselves in this position). We might also recall Alison's descent down the flight of stairs in Problem 2.1, where knowledge that the Fibonacci numbers often appear in these kinds of problems made our initial conjecture more enticing after checking only a few values.

This illustrates how individuals with different levels of knowledge and experience may react to the same quasi-empirical evidence differently. This may occur even if they are otherwise similar in their goals and intentions, psychological make-up, belief-forming mechanisms, and attitudes towards risk. Beyond these psychological questions of influence, our background knowledge can also affect how rational it is to believe a conjecture on the basis of a given body of evidence. The non-deductive evidence will have no effect on us if we already know a proof or a counterexample. This given, such methods are clearly less suitable for

maintaining Consensus amongst a diverse body of mathematical researchers with different backgrounds, who are thus likely to diverge in their judgements.

Lastly, even when we have explicitly reflected on the truth of the two auxiliary results some sliver of doubt about our conjecture may remain, and although it seems that it would take a particularly stubborn person to continue to withhold assent at this point, we would not know how to go about convincing someone who was still left unsatisfied. But how elegantly these anxieties are alleviated by discovering a proof:

Let $f : (0, \frac{\pi}{3}] \to \mathbb{R}$ be given by $f(x) = \frac{1}{\frac{4\pi}{x}\tan(x)}$.

Then we have that

$$f'(x) = \frac{d}{dx}\left(\frac{1}{\frac{4\pi}{x}\tan(x)}\right) = \frac{1}{4\pi}\frac{d}{dx}\left(\frac{x\cos(x)}{\sin(x)}\right)$$

$$= \frac{1}{4\pi}\left(\frac{(\cos(x) - x\sin(x))\sin(x) - x\cos^2(x)}{\sin^2(x)}\right)$$

$$= \frac{1}{4\pi}\left(\frac{\cos(x)\sin(x) - x}{\sin^2(x)}\right)$$

$$< \frac{1}{4\pi}\left(\frac{\sin(x) - x}{\sin^2(x)}\right) < 0$$

as $sin(x) < x$ for $x \in (0, \frac{\pi}{3}]$.

This means that $f(x)$ is a decreasing function of x. Switching to a new variable $t = \pi/x$ we get a function $g: [3, \infty) \to \mathbb{R}$ which is now strictly increasing, so if $t_1 > t_2$ then $g(t_1) > g(t_2)$. But $g(n) = p_n$ for all integers $n \geq 3$, and so the truth of Conjecture 2.5 follows. This argument will be found convincing in and of itself by any competent reader, independently of their background knowledge and experience. It is thus far more effective for establishing Consensus that the result is true, and there can no longer be rational concern about its Reliability.

2.v. The Goldbach Conjecture

Our next example is a claim of some historical significance, which has fascinated mathematicians for centuries: the Goldbach Conjecture. We will see that although some specialist mathematicians have yielded Private Acceptance to the conjecture on the basis of non-deductive evidence alone, it is not clear whether such inferences are sufficiently Reliable. Moreover, these arguments for the truth of the

claim are in some degree dependent upon the authority of the specialists making the judgement, undermining Autonomy. We will thus see that Public Acceptance of the result is not yet appropriate – as has been born out in practice.

In a letter to Christian Goldbach, dated 30th June 1742, Euler writes:

> '… every even integer is a sum of two primes. I regard this as a completely certain theorem, although I cannot prove it.'[84]

This is the famous Goldbach Conjecture (herein 'GC'). Euler was following a since-abandoned convention that 1 is a prime: for us the conjecture pertains only to even integers greater than 2.

Conjecture 2.8. (Goldbach Conjecture)

Let n be an even integer greater than 2. Then n can be expressed as the sum of two (not necessarily distinct) primes.

Since its emergence from the Goldbach-Euler correspondence, GC has been subject to quasi-empirical investigation on a massive scale. Several mathematicians have checked that it is not falsified by large initial segments of the natural numbers, and in 2013 a lower bound of 4×10^{18} for any counterexample was given.[85] Despite all of this inductive evidence it seems that a proof is not forthcoming. Number theorist and Fields medalist Alan Baker stated in a 2000 interview that 'It is unlikely that we will get any further without a big breakthrough. Unfortunately there isn't such a big idea on the horizon.'[86]

As already mentioned earlier, some mathematicians have given full Private Acceptance to GC, notwithstanding the lack of a proof, and not only Leonhard Euler. Indeed, Echeverria has claimed that 'the certainty of mathematicians about the truth of GC is complete'.[87] And in 1922 the eminent number theorists John Littlewood and Godfrey Hardy were willing to assert that 'there is no reasonable doubt that the theorem is correct',[88] all before the conjecture had even been checked up to 10^5. GC thus provides another example of the extension of non-deductive methods to the context of justification – one endorsed by a substantial number of mathematicians.

---

[84] Władysław Narkiewicz, *The Development of Prime Number Theory: From Euclid to Hardy and Littlewood* (New York: Springer-Verlag, 2000), 333.

[85] Tomás Oliveira e Silva, Siegfried Herzog, and Silvio Pardi, "Empirical Verification of the Even Goldbach Conjecture and Computation of Prime Gaps up to 4×10^{18} ", *Mathematics of Computation* 83 (2013): 2033-2060.

[86] Anjana Ahuja, "A million-dollar maths question", *The Times,* March 16, 2000).

[87] Javier Echeverria, "Empirical Methods in Mathematics. A Case-Study: Goldbach's Conjecture", in *Spanish Studies in the Philosophy of Science,* ed. G. Munévar (Boston: Kluwer Academic Publishers, 1996), 19-55.

[88] Quoted in Alan Baker, "Non-deductive methods in mathematics", *Stanford Encyclopedia of Philosophy*, accessed 10th August 2015, http://plato.stanford.edu/entries/mathematics-nondeductive/ By coincidence, the author of the article has the same name as the mathematician quoted.

Psychologically speaking, the data gathered by Tomás and others do seem convincing. Let us then consider how we might work these data into the presentation of an argument that establishes the truth of GC. We know that an admittedly huge initial run of the even natural numbers have been checked and are consistent with GC. What general principle could be invoked to move us from these data to the truth of the conjecture itself? One line of thought which might come naturally here is that if there were a special property belonging to some integers, possession of which by an integer entailed that it could not be split into the sum of two primes, then integers with this property would have already been encountered by now: surely a sample of size 4×10^{18} is sufficiently representative. However, consider the following problem.[89]

Problem 2.9.

Is it true that $991n^2 + 1$ is never a perfect square?

Suppose that we had calculated this expression for a huge initial segment of the natural numbers, well beyond Tomás' investigation for GC – every value up to 10^{25}, say – and the expression was never a square. According to the same reasoning as above, this would suffice to demonstrate the conjecture that no such numbers exist. But this conjecture is actually false: the first counterexample is $n = 12{,}055{,}735{,}790{,}331{,}359{,}447{,}442{,}538{,}767 \approx 1.2\times10^{28}$. Moreover, this is no isolated example: historically there have been many other plausible-seeming claims in number theory that have turned out to have only huge smallest counterexamples. Here are three others:

- In 1769, Euler conjectured that for all integers n and k greater than 1, if the sum of n k^{th} powers of non-zero integers is itself a k^{th} power, then n is greater than or equal to k.[90] A counterexample for $k = 5$ was found[91] in 1966: it is now easy to verify via computer that $144^5 = 27^5 + 84^5 + 110^5 + 135^5$. A counterexample[92] for $k = 4$ is $422481^4 = 95800^4 + 217519^4 + 414560^4$

- In 1885, Thomas Joannes Stieltjes conjectured in a letter to Hermite that the following claim was true.[93] Define the Mertens function as

[89] Adapted from Joseph Rotman, *Journey into Mathematics: An Introduction to Proofs* (New Jersey: Prentice Hall, 1998), 20.
[90] Leonard Eugene Dickson, *History of the Theory of Numbers*, *Vol. 2* (Chelsea: New York, 1952), 658.
[91] L. J. Lander and T. R. Parkin, "Counterexample to Euler's Conjecture on Sums of Like Powers", *Bulletin of the American Mathematical Society* 72 (1966): 1079.
[92] Roger Frye, "Finding $95800^4 + 217519^4 + 414560^4 = 422481^4$ on the Connection Machine", *Proceedings of Supercomputing* 88 (1988): 106-116.
[93] Reprinted in Thomas Stieltjes, "Lettre a Hermite de 11 Juillet 1885" in B. Bailaud and H. Bourget, *Correspondance of d'Hermite et Stieltjes* (Paris: Gauthier-Villars, 1905), 160-164.

$M(n) = \sum_{1 \le k \le n} \mu(k)$, where $\mu(n)$ is the Möbius function.[94] Then for all $n > 1$, $|M(n)| < \sqrt{n}$. This conjecture is known to imply the Riemann Hypothesis, but in 1985 the existence of a counterexample between 10^{14} and $e^{3.21 \times 10^{64}}$ was proved by Andrew Odlyzko and Herman te Riele.[95]

- In 1919, George Pólya conjectured that at least half of the natural numbers less than any given natural number n have an odd number of prime factors when counted with multiplicity.[96] The conjecture was disproved by Colin Brian Haselgrove in 1958.[97] The first explicit counterexample $n = 906{,}180{,}359$ was given by Russell Sherman Lehman in 1960.[98] The smallest counterexample is $n = 906{,}150{,}257$, found by Minoru Tanaka in 1980.[99]

There are also very large numbers used in mathematical proofs: for example, Graham's number and Moser's number, which require special notation to express. Presumably these numbers are the smallest we are aware of having certain interesting properties. Some of these conjectures or others like them could be relevant to the truth of GC: the suggestion is at least not obviously implausible.

These kinds of examples raise a concern about the Reliability of believing the conjecture on this basis alone and make it seem unlikely that there is any general principle that enables us to complete the argument. Indeed, because number-theoretic properties of the integers are not in general uniformly distributed – the prime numbers have asymptotically zero density, for example – they are not conducive to this kind of treatment, and there is *prima facie* no special reason to assume the property of satisfying GC to be an exception. We can find arbitrarily long sequences of consecutive integers with no primes: $(n+1)!+2, (n+1)!+3, \ldots, (n+1)!+(n+1)$ is a run of n composite numbers for any natural number n. Yet no conclusion about the non-existence of prime numbers follows. But there might not be anything particularly special or indicative about the *first* run of 4×10^{18} integers with respect to GC; our data may be highly biased.

One potential objection to what I have said so far is to note that the account I have given of the evidence is unfair. Confidence in the result actually only increased when it was pointed out that the *number* of ways of expressing n as the sum of two primes appears to increase with n. Consider the following table of $G(n)$, the number of Goldbach partitions of n, for a few small values of n.

---

[94] I.e. $\mu(n)$=0 if n is a multiple of a square number other than 1, $\mu(n) = 1$ if n is square-free with an even number of prime factors, and $\mu(n) = -1$ if n is square-free with an odd number of prime factors.

[95] Andrew Odlyzko and Herman te Riele, "Disproof of the Mertens Conjecture", *Journal für die reine und angewandte Mathematik* 357 (1985): 138-160.

[96] George Pólya, "Verschiedene Bemerkungen zur Zahlentheorie", *Jahresber Deutschen Math.-Verein* 28 (1919): 31-40.

[97] Colin Haselgrove, "A Disproof of a Conjecture of Pólya", *Mathematika* 5 (1958), 141-145.

[98] Russell Lehman, "On Liouville's Function", *Mathematics of Computation* 14 (1960): 311-320.

[99] Minoru Tanaka, "A Numerical Investigation on Cumulative Sum of the Liouville Function", *Tokyo Journal of Mathematics* 3 (1980): 187-189.

n	$G(n)$	n	$G(n)$	n	$G(n)$	n	$G(n)$
4	1	32	2	60	6	88	4
6	1	34	4	62	3	90	9
8	1	36	4	64	5	92	4
10	2	38	2	66	6	94	5
12	1	40	3	68	2	96	7
14	2	42	4	70	5	98	3
16	2	44	3	72	6	100	6
18	2	46	4	74	5	102	8
20	2	48	4	76	5	104	5
22	3	50	4	78	7	106	6
24	3	52	3	80	4	108	8
26	3	54	4	82	5	110	6
28	2	56	3	84	8	112	7
30	3	58	4	86	5	114	1 0

Moreover, if we plot $G(n)$ for the first 10^6 values of n a beautiful pattern emerges:

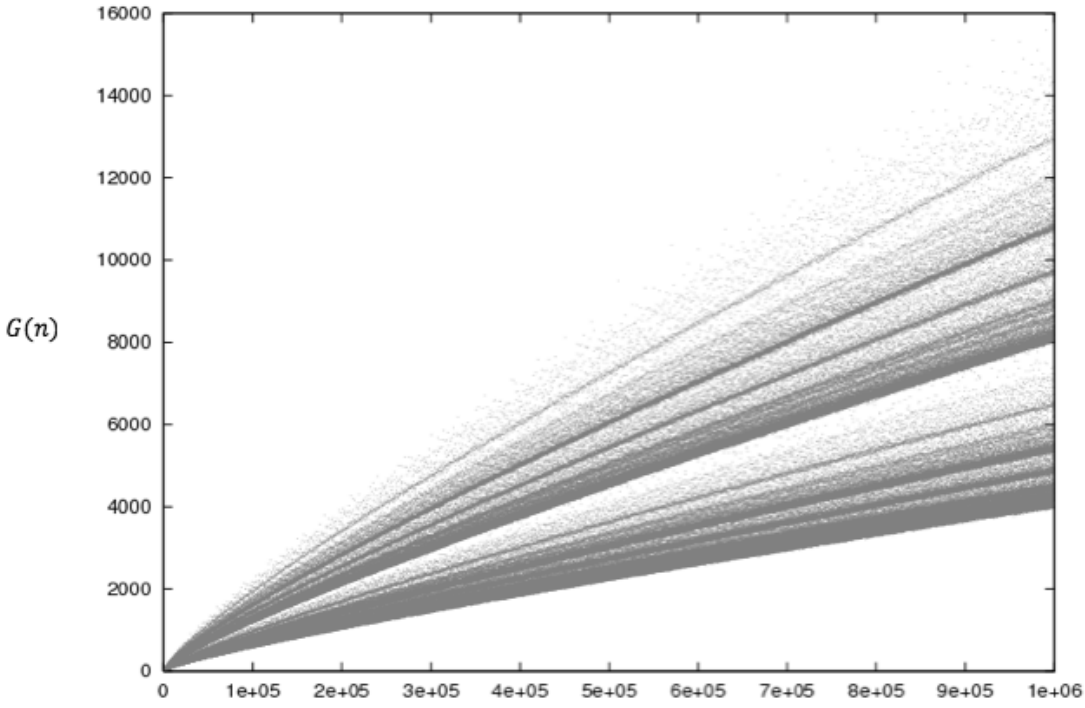

Figure 2.9. A graph of G(n) for the first million values of n. '$ae+0b$' means $a\times10^b$. http://www.cs.mcgill.ca

The strange regularity of the pattern does add some strength to the claim that $G(n)$ tends to get bigger with n. But this evidence is really no more conclusive, despite the richness and psychological power of the visual pattern. It is not clear that $G(n)$'s being large mitigates the possibility that $G(n+1) = 0$, and the claim that this rather pleasing pattern will continue is itself only supported inductively. It appears, then, that this additional evidence will be of no help in reconstructing a convincing argument for GC. Yet the opinions of Littlewood and Hardy on matters concerning number theory are not to be dismissed lightly. How best can we account for their confidence?

As an attempt at a complete elucidation of the grounds upon which Littlewood, Hardy and other mathematicians have come to believe GC, the presentation I have given so far is indeed highly misleading, and indeed necessarily so. For Littlewood and Hardy's confidence will not have been grounded in explicit inference from inductive evidence based on some general theoretical principle, or perhaps some variant of inductive logic. Rather, they will have come to an intuitive judgement about the plausibility of the conjecture in light of all the available evidence they possessed. And as the discussion in the previous section illustrates, this judgement will itself have been informed and influenced by many other things they had known, believed and experienced up to that point in their mathematical careers. Likewise for any contemporary number theorist who now believes the truth of GC. We give three examples of related theorems, prior knowledge of which may have affected our assessment of the import of the quasi-empirical data.

Theorem 2.10. (Ternary Goldbach Conjecture)

Let n be an odd integer with $n > 5$. Then n is the sum of three (not necessarily distinct) primes.

The Ternary Goldbach Conjecture was stated by Goldbach himself in an earlier letter to Euler.[100] It is sometimes also known as the Weak Goldbach Conjecture, because it is a simple corollary of GC, as Euler noted (e.g., if n is odd and $n > 5$, and if GC is true, then $(n-3)$ must be the sum of two primes). Although the proof the theorem was not completed until the efforts of the Peruvian mathematician Harald Helfgott in 2013,[101] Littlewood and Hardy were themselves able to show in 1923 that given the generalized Riemann Hypothesis the conjecture was true for sufficiently large numbers,[102] and it is likely that this achievement would have not only increased their confidence in the ternary conjecture, but also in the stronger result GC itself.

Theorem 2.11.

Let n be an even integer greater than 2. Then n is the sum of a prime and a second number that is the product of at most two primes.

---

[100] Leonard Eugene Dickson, *History of the Theory of Numbers*, *Vol. 1* (Chelsea: New York, 1952)
[101] Harald Helfgott, "The Ternary Goldbach Conjecture is True", 2013. arXiv:1312.7748
[102] Godfrey H. Hardy and John E. Littlewood, "On Some Problems of "Partitio Numerorum" III: On the Expression of a Number as the Sum of Primes", *Acta Mathematica* 44 (1923): 1-70.

Clearly, this result by Chen Jingrun is progress from another direction: it seems that mathematicians are homing in on the full result GC.[103]

Theorem 2.12.[104]

The set of even integers that cannot be represented as the sum of two primes has asymptotic density 0.

This claim bears directly on the import of the non-inductive evidence: it seems to makes it more reasonable to attempt the inductive strategy, because counterexamples must eventually become very sparse over a long enough initial segment. Yet its exact implications remain unclear: have we checked a sufficient number of cases for the evidence to now be compelling, or will the asymptotic density condition become effective only much later on?

Viewing mathematicians' attitude towards non-deductive evidence as stemming from intuitive judgements based on professional experience means that we need not commit them to any dubious general principle about all number theoretic claims inviting belief when they are verified for a large enough initial segment of the natural numbers. Yet it also seems to preclude the presentation of an explicit argument that is compelling unto itself. It is impossible for the whole evidential and psychological history of an intuitive judgement to be externalised from the mathematician making it and published alongside the conjecture. The Public Acceptance of GC on this basis would therefore undermine Autonomy.

We have also seen a glimpse of the wide range of different factors that might affect our intuitive judgements when a sizeable number of researchers direct their attention to a problem. So again there may be issues with Consensus as different experts weigh in about how convincing the data are. There will indeed always be further insights available – be they related conjectures, historical examples, or theoretical considerations. And discovery of each of these may cause judgements about the evidence to shift back and forth over time, also damaging Permanence. It seems then that mathematicians are rationally justified in withholding Public Acceptance of GC, even though it may be widely Privately Accepted by mathematicians.

2.vi. Acceptance Without Proof

In this chapter, we have seen that non-deductive evidence can give psychologically compelling grounds for believing a mathematical claim. In the last section, we also saw that mathematicians themselves have yielded Private Acceptance to GC on these kinds of grounds alone. Consider another historical

[103] Chen Jingrun, "On the Representation of a Large Even Integer as the Sum of a Prime and the Product of at Most Two Primes", *Kexue Tongbao* 11 (1966): 385-386.
[104] Carl Pomerance and Richard Crandal, *Prime Numbers: A Computational Perspective* (New York: Springer, 2001), 17.

example. In his *Opera Omnia*, Euler conjectures that the divisor function[105] $\sigma(n)$ satisfies the following recurrence relation:[106]

$$\sigma(n) = \sigma(n-1) + \sigma(n-2) - \sigma(n-5) - \sigma(n-7) + \sigma(n-12) + \sigma(n-15) - \ldots$$

where the signs alternate in twos. The sum continues until the arguments for σ become negative, and he asks us to define $\sigma(0)$ as equal to n if this expression occurs in the sum. The sequence 1,2,5,7,12,15, ... of numbers subtracted from n is given recursively: the differences between successive terms are $\mathbf{1}, 3, \mathbf{2}, 5, \mathbf{3}, 7, \mathbf{4}, 9, \mathbf{5}, \ldots$ i.e. an alternation of the natural numbers and the odd numbers starting with 3.

After explaining the content of his conjecture, Euler verifies the formula for $n = 1$ through to $n = 20$. He then writes, 'I think these examples are sufficient to discourage anyone from imagining that it is by mere chance that my rule is in agreement with the truth'.[107] After considering that he might have given the wrong formula for the series of subtracted numbers, he uses his claim to correctly predict that $\sigma(101) = 102$ and $\sigma(301) = 352$, and then comments that 'The examples I have just developed will undoubtedly dispel any qualms which we might have had about the truth of my formula.'[108] Euler later considers a related result:

$$(1-x)(1-x^2)(1-x^3)(1-x^4)\ldots = 1 - x - x^2 + x^5 + x^7 - x^{12} - x^{15} + \ldots$$

where the patterns of minus signs and exponents of *x* are the same as in the recurrence relation given above. This time he writes that in order that to establish this result 'It suffices to undertake this multiplication and to continue it as far as it is deemed proper to become convinced of the truth of this series.'[109] He goes on: 'I have proposed the same question to some of my friends with whose ability in these matters I am familiar, but all have agreed with me on the truth of this transformation of the product into a series, without being able to unearth any clue of a demonstration.'[110] Euler then describes the result as a 'truth' that is 'known' but 'not yet demonstrated'.[111]

This kind of Private Acceptance in the absence of proof is in fact fairly common, both historically and in contemporary mathematics. Consider now the remarks of the Hungarian-born American mathematician Paul Halmos, who in a 1984 article gave a description of his experiences of what working as a practicing mathematician is actually like:

---

[105] $\sigma(n)$ is defined for natural numbers n as the sum of the (positive) divisors of n, so that $\sigma(7) = 8$ and $\sigma(12) = 28$

[106] Leonhard Euler, *Opera Omnia*. Quoted in George Pólya, *Mathematics and Plausible Reasoning, Volume I* (Princeton: Princeton University Press, 1954), 92-94.

[107] Ibid., 94.

[108] Ibid., 95.

[109] Ibid., 96.

[110] Ibid., 100.

[111] Ibid., 100.

> Mathematics – this may surprise you or shock you some – is never deductive in its creation. The mathematician at work makes vague guesses, visualizes broad generalizations, and jumps to unwarranted conclusions. He arranges and rearranges his ideas, and he becomes convinced of their truth long before he can write down a logical proof.[112]

After quoting Halmos, De Villiers goes on to write that 'a very high level of conviction may sometimes be reached even in the absence of a proof'. Yet despite this importance in generating Private Acceptance and in guiding research more generally, there has always been a strong feeling amongst mathematicians that non-deductive arguments are never a sufficient basis for the Public Acceptance of new results. (This is of course a principle consequent upon the insistence on proof for Public Acceptance discussed in Chapter 1.) Consider Euler again, later on in a passage quoted earlier (Section 2.i):

> "The kind of knowledge which is supported only by observations and is not yet proved must be carefully distinguished from the truth; it is gained by induction, as we usually say. Yet we have seen cases in which mere induction led to error. Therefore, we should take great care not to accept as true such properties of the numbers which we have discovered by observation and which are supported by induction alone."[113]

To make sense of his other remarks, and his use of (a term translated by) 'knowledge', it is reasonable to assume that Euler's phrase 'accept as true' means something like Public Acceptance rather than Private Acceptance here.

Since Euler's time, however, we have seen that the development of the personal computer has increased the potential scope of non-deductive methods immensely. Consequently, there has been some suggestion that the long-upheld rule of proof for Public Acceptance should be revised to accommodate new routes to justification. De Villiers discusses the views of the mathematician Branko Grünbaum, who used the computer application *Mathematica* to 'explore and verify' some geometric results:

> Do we start trusting numerical evidence (or other evidence produced by computers) as proofs of mathematics theorems? ... if we have no doubt—do we call it a theorem? ... I do think my assertions are theorems ... the mathematical community needs to come to grips with new modes of investigation that have been opened up by computers.[114]

In the remainder of this section I will use the insights from the examples discussed earlier in the chapter to argue that there are good reasons for mathematicians to continue to reject at least the majority of these methods – with the possible exception of some probabilistic algorithms – as a basis for the Public Acceptance of results. The argument will proceed through consideration of our four Practical

---

[112] Paul Halmos, "Mathematics as a creative art", in *Mathematics: People, Problems, Results, Vol 2* (California: Wadsworth, 1984), 23

[113] Quoted in George Pólya, *Mathematics and Plausible Reasoning, Volume I* (Princeton: Princeton University Press, 1954), 3.

[114] Branko Grünbaum, "Quadrangles, Pentagons, and Computers", *Geombinatorics* 3 (1993): 8.

Virtues of mathematical practice: Permanence, Reliability, Consensus, and Autonomy.

As Euler points out, many conjectures for which apparently convincing evidence is available have turned out to be false. If these conjectures had been Publicly Accepted, then clearly the mathematical literature would have become less Reliable as a result, and Permanence may have suffered too. If journals are to permit mathematicians to unqualifiedly assert results on the basis of non-deductive methods alone, it is therefore important that these methods are known to be adequately reliable guides to the truth of the results they are being used to support.

Just as proof presentations must be found acceptable by the entire readership of a journal, so too must other mathematicians agree that the methods used to justify a published claim are indeed Reliable. Yet as we saw in the last two sections, both experience and background knowledge can affect intuitive judgements about whether a body of evidence warrants regarding a result as established. Such judgements are therefore unlikely to be unequivocal, even once all the available information is taken into account. In this chapter I have in several places expressed a hope that my intuitive views about the different examples would be shared; yet the reader may have found himself or herself disagreeing in places. So if assessment of whether a given method is acceptable was to be carried out on an intuitive and unsystematic basis, it is likely that Consensus would soon fragment.

Similarly, as new evidence appears and is taken into account, individual mathematicians may change their minds about whether a conjecture has been adequately supported as time goes on. The Permanent character of mathematics would thus be eroded if results were to be Publically Accepted on such as basis.

A related point concerns Autonomy. Consider, for example, the claim of Davis and Hersh that the non-deductive evidence available for the Riemann Hypothesis is 'so strong that it carries conviction even without rigorous proof'.[115] Perhaps they are warranted in making this claim, but how could a non-specialist decide either way? Whether intuitive judgements are justified in a particular case is never ascertainable by an external observer, because the ultimate evidential, experiential and psychological basis of the judgement is always ultimately inscrutable. Rigorous proofs presentations, on the other hand, can be checked by any competent mathematician, even if they are not a specialist in that area of mathematics and could not have come up with the argument themselves. Compare this with chess: I can perhaps read about and understand a match played by Kasparov or Fischer, despite being unable to originate the same moves myself; certainly I can check that all the moves were legal.

Consider again Euler's expansion of the product into a series given above, and the clear appeal to authority his remarks contain. The question of how far it is 'deemed proper' to continue working out terms of the series is not answered objectively or systematically. Rather, he justifies his own stopping point by reporting an agreement between the intuitive judgements of his friends, whom he

[115] Philip Davis, Reuben Hersh and Elena Anne Marchisotto, *The Mathematical Experience: Study Edition* (New York: Berkhäuser, 2012), 411.

states are sufficiently mathematically able to count as qualifying as competent judges, a claim which again rests only upon his own testimony.

In these kinds of cases, the explicit mathematical justification given is not in and of itself sufficient to compel assent, and the arguments are completed only through the intuitive judgements of the specialist, in whose authority the justification for the claim then partly rests. This may create practical problems in educational contexts: suppose we have a student who is particularly recalcitrant and refuses to be convinced by any amount of quasi-empirical evidence. We would then have no further recourse than asserting our authority. However, if we can successfully show them that a proof is available then this will rationally compel assent whether the student wishes to co-operate or not (again c.f. Tall).

What would be needed in order to preserve the four Practical Virtues is a shared, systematic way of telling whether a given body of non-empirical evidence is sufficient to warrant the acceptance of a conjecture. These techniques must be sufficiently stringent to remove any issues about Reliability, and the verdicts they render must be stable over time, so as to maintain Permanence. To preserve Autonomy the evaluative techniques must be such that any competent mathematician can apply them: this will require them to be fully transparent and given by a definite procedure. Lastly, in order to be suitably precise and to maintain Consensus, the verdict rendered must be unequivocal. If grounded in natural language they will lack the necessary precision and objectivity. The outcome will therefore need to take a numerical value. Such quantitative relations between evidence and conjecture are the province of the theory of probability.

It is also clear that the existence of techniques meeting all of these requirements will only be possible if our non-deductive methods themselves are delineated sufficiently sharply. Otherwise, it is impossible to be systematic and we are back to the kind of skill embedded in specialists' intuition. Yet so far we have encountered mainly opportunistic investigations that yielded data that were construed as evidence only in hindsight: not really *methods* as such, in the sense of a definite procedure that can be applied to many problems. In the coming chapters, then, we will focus our attention on definite algorithms whose reliability can be determinately evaluated using probabilistic techniques.

2.vii. Conclusion

In this chapter, we have examined a number of examples of the use of non-deductive methods in the context of justification. Issues were raised about the epistemic security of these techniques, about the unequivocality of judgements that are often sensitive to experience and background knowledge, and the need for reliance on the judgement of specialists that a given body of evidence is sufficient to warrant assent. Where these reservations apply, non-deductive methods are always unsuitable for justifying mathematical claims in the context of Public Acceptance, on pain of the rapid deterioration of Permanence, Reliability, Consensus and Autonomy – though individual mathematicians may in some instances be warranted in Privately Accepting claims on this basis alone.

I have also argued that the only kind of non-deductive methods which might permit the preservation of the Practical Virtues when employed in the context of Public Acceptance are probabilistic algorithms: definite procedures for which a provable, quantitative assessment of reliability is available. Happily, in the next chapter we shall see that such methods do in fact exist and find application in many different areas of mathematics.

3. Randomised Algorithms

This chapter discusses the growing class of computer-based mathematical algorithms that make essential use of randomness. In the cases we are most interested in, the algorithms do not guarantee us the correct answers, giving us strong but less than conclusive reasons for accepting the solutions they provide. Our focus will be further restricted to algorithms that enable us to give a quantitative, objective evaluation of the strength of the evidence they supply: precisely what was found crucially lacking for the non-deductive methods presented in the previous chapter. The most important of these for the present enquiry will be a version of the Rabin-Miller Algorithm that is very likely to produce a prime number, with a probability of failure that can be precisely bounded, and in principle made smaller than any fixed positive number.

3.i. Las Vegas and Monte Carlo Algorithms

We begin with a brief introduction to randomised algorithms. Various precisifications of the term 'algorithm' have been attempted, but for our purposes it will suffice to say that an algorithm is a finite ordered list of discrete instructions, which collectively provide an effective means to achieving some specific goal.[116] Each instruction must be clear enough for us to be able to tell when we have completed it, and its completion must always be possible. We restrict discussion to algorithms for carrying out some distinctly mathematical task, such as multiplying two natural numbers, solving an algebraic equation, finding a prime number, or extracting a root. For these mathematical algorithms, the instructions will be given in terms that are explicit enough to be conducive to implementation on a computer.

The word 'algorithm' comes from *Algoritmi*, the Latinised form of the name al-Khwārizmī, a ninth century Persian mathematician who wrote treatises giving rules for performing arithmetic operations on integers using the Hindu Numeral system, and for solving linear and quadratic equations. In this spirit, we begin with a presentation of the modern 'completing the square' algorithm for solving quadratic equations. Instructions are accompanied by a demonstration with a concrete example.

Algorithm 3.1. Completing the Square

1. Collect all the terms to one side, writing higher powers of x first:

$$3x^2 - 42x + 144 = 0$$

[116] Attempts to characterise algorithms in more detail have been made by Chomsky, Kleene, Turing, Gödel, Minsky and others.

2. Divide each term through by the x^2 coefficient:

$$x^2 - 14x + 48 = 0$$

3. Subtract the constant term from both sides, leaving only x terms on the left-hand side:

$$x^2 - 14x = -48$$

4. Add the square of half of the x coefficient to both sides:

$$x^2 - 14x + 49 = 1$$

5. Write the left-hand side as the square of a linear expression in x.

$$(x - 7)^2 = 1$$

6. Take square roots of both sides, allowing for the negative root on the right-hand side.

$$x - 7 = \pm 1$$

7. Subtract the remaining constant term on the left-hand side from both sides:

$$x = 7 \pm 1 = 6 \; or \; 8$$

Here the requisite precision has been achieved using a combination of explicit instructions and illustration. It is a platitude amongst teachers of mathematics that practical competence in such techniques is best achieved using repetition with numerical examples in addition to explicit elucidation. Computers, on the other hand, can follow any coherent and sufficiently precise set of instructions we give, and to do so perfectly the first time round. However, when written in code, the steps must be in a specific format that is more explicit than the natural language description given above.

In implementing an algorithm on a computer, we program that computer in such a way that when an input of specified form is given it performs a series of determinate calculations corresponding to the algorithm. This yields an output that is displayed for the user to see. For instance, when we input coefficients (a, b, c) a suitably programmed computer might then calculate the solutions of the corresponding quadratic equation, or inform us of the number of distinct real solutions. An algorithm in this sense thus constitutes a definite method for solving a specific class of mathematical problems.

We define a randomised algorithm as one where at least one instruction is of the form 'Select a random entry from the set S'. The selection is assumed to be unbiased, in that each member of S has an equal chance of being selected, and

each selection is independent. We will also assume that S is finite.[117] As well as an initial input given explicitly by the user, randomised algorithms may therefore require these randomly selected inputs to be supplied at any time during their implementation.

In a pioneering 1976 paper, the mathematician Michael O. Rabin – one of the most important exponents of randomised algorithms – gave some of the first concrete examples of randomised algorithms in computational geometry and number theory, as well as clarifying some of the underlying concepts. One of these algorithms was designed to find solutions to the closest pair problem; that is, given a set S of n points in a metric space, to find a pair of minimal distance apart.[118] Many variants of this algorithm now exist, some of which are now able to keep track of the changing solutions as points are added and subtracted from S.[119]

Interest in randomised algorithms has increased tremendously in recent years. Such algorithms are now available for many tasks in diverse areas of pure and applied mathematics, such as number theory, combinatorics, graph theory, sorting problems, numerical analysis, statistical physics, and in performing simulations.[120] Motwani and Raghaven suggest two reasons for this surge in interest: randomised algorithms are often much faster than their deterministic equivalents, and typically have a very simple structure that is often surprisingly easy to describe and implement.[121]

This dramatic increase in speed and simplicity is of interest to us because of our observation in the first chapter that the growing length of mathematical argumentation has been in some respects problematic. If their applicability is sufficiently broad, perhaps these randomised algorithms can help mathematicians to maintain the four Practical Virtues of mathematical enquiry. Let us now consider a concrete example of a randomised algorithm.

In 1959, the British computer scientist Tony Hoare discovered a way of sorting an ordered set $S = (x_1, x_2, \dots, x_n)$ of n real numbers into ascending order.[122] His solution to the problem used recursion. We select an element $x \in S$, called the 'pivot'. We then Partition $S/\{x\}$ into two ordered subsets S_1 and S_2 such that x is greater than each element of S_1, but less than or equal to each element of S_2. Clearly, if we can order both smaller subsets S_1 and S_2 then our task is complete. So we attempt to sort S_1 and S_2 using the same method. Because the subsets get

---

[117] The precise meaning of this random selection will be explored in great detail in chapter 6 (see Sections 6.ii–6.v).

[118] Michael Rabin, "Probabilistic Algorithms", in *Algorithms and Complexity: New Directions and Recent Trends*, ed. J. F Traub (New York: Academic Press, 1976), 21-39.

[119] Mordecai Golin, Rajeev Raman, Christian Schwarz, and Michiel Smid, "Randomized Data Structure for the Dynamic Closest Pair Problem", *Society for Industrial and Applied Mathematics*, 27 (1998): 1036-1072.

[120] For a survey, see Richard M. Karp, "An Introduction to Randomized Algorithms", *Discrete Applied Mathematics* 34 (1991): 165-201. See also See Rajeev Motwani and Prabhakar Raghaven, *Randomized Algorithms*, (Cambridge: Cambridge University Press, 1995), ix, 23-24.

[121] Motwani and Raghaven, *Randomized Algorithms*, ix, 6-7. Superiority can be provable: see ibid. 74.

[122] Sir Charles 'Tony' Hoare, "Algorithm 64: Quicksort", *Communications of the ACM* 4 (1961): 321.

strictly smaller each time, eventually we get to subsets of size 1 or 0, which obviously don't need sorting. So as long as we keep track of the pivots and how the members of the smaller subsets are ordered in relation to them, we are done.

A further question remains before we can implement this algorithm: how do we choose the pivot? One idea is to always select the element from a specific place in the ordered set S, such as the first entry. However, this procedure yields a poor choice when the list is already sorted, which may happen frequently in practice. A better strategy may be to select the pivot randomly from S.[123]

Algorithms such as Randomized Quicksort that in principle always yield a correct output but may have unpredictable running times are called 'Las Vegas' algorithms. The term was introduced by László Babi in 1979, in a paper about finding colour-preserving isomorphisms between vertex-coloured graphs.[124] Most authors also include the condition that the expected running time – taken over all possible values of the randomised input – must be finite for each initial input. Other early examples of Las Vegas algorithms include the algorithms for solving the closest pair problem mentioned above, Berlekamp's algorithm for factoring polynomials over large finite fields,[125] randomised methods for finding square roots of integers modulo a prime p,[126] Zippel's algorithm for finding the GCD of two polynomials,[127] and algorithms for finding a cut of minimal cardinality in an undirected multigraph.[128]

Another important subclass of these methods is Monte Carlo algorithms. A randomised algorithm is Monte Carlo if it sometimes gives an incorrect solution to the task it is designed to solve.[129] Following the instructions exactly is thus not a logically sufficient guarantee of completing the task, and a list of completed operations corresponding to the performance of a Monte Carlo algorithm is never isomorphic to a proof that a correct solution has been found. Unlike Las Vegas algorithms, Monte Carlo algorithms therefore never yield grounds regarded as sufficient for the Public Acceptance of the solutions they provide. Examples of Monte Carlo algorithms include further procedures for finding graph isomorphisms,[130] and a randomized version of the Schreier-Simms Algorithm for computing a base[131] and a strong generating set[132] for a finite permutation group.

---

[123] Motwani and Raghavan, *Randomized Algorithms*, 3-4.
[124] László Babai "Monte-Carlo Algorithms in Graph Isomorphism Testing", *Technical Report of the DMS* 79 (1979).
[125] Elwyn Berlekamp, "Factoring Polynomials over large finite fields", *Mathematics of Computation* 24 (1970): 713-735.
[126] Crandal and Pomerance, *Prime Numbers*, 93-94.
[127] Richard Zippel, "Probabilistic Algorithms for Sparse Polynomials", in *Symbolic and Algebraic Computation*, ed. Edward Ng (New York: Springer, 1979), 216-226.
[128] Motwani and Raghavan, *Randomized Algorithms*, 8. A multigraph is a graph that may have more than one edges between any pair of vertices. A cut is a set of edges who removal will split a given undirected graph or multigraph into two or more components.
[129] A Monte Carlo algorithm may have a determinate running time and so not be Las Vegas – for instance, the Rabin-Miller Algorithm given below.
[130] Babai, "Monte-Carlo Algorithms in Graph Isomorphism Testing."
[131] A base for a permutation group G acting on a set X is a set $B \subset X$ such that only the identity element $e_G \in G$ fixes every element of B.
[132] For a definition, see Ákos Seress, *Permutation Group Algorithms* (Cambridge: Cambridge University Press, 2003), 55.

This calculation is important because it allows swift computation of the order and other properties of the group, as well as determination of group membership.[133] As we shall see later with the Rabin-Miller Algorithm, there are also important applications of Monte Carlo algorithms in finding large prime numbers.

The algorithms considered in the rest of this chapter all have a certain kind of structure, which we will call Iterative Monte Carlo (IMC) algorithms. An IMC algorithm takes some mathematical object x and attempts to determine whether or not it has a given property P. For instance, an IMC might complete a task such as 'Given a number n, determine if n is prime' or 'Given a graph G, determine if G is planar'. It is therefore an algorithm for solving a decision problem: that is, a problem for which the answer is 'yes' or 'no'. Furthermore, the operation of an IMC will consist in running a finite number of identical tests, each of which has two outcomes: 'pass' or 'fail'. The IMC outputs 'no' if the object x fails any of the tests, and 'yes' if and only if it passes all of them.

The IMC tests for the algorithms given in the next few sections all work as follows. Firstly, given our object x, there will be a set S_x associated with x, such that if x in fact lacks the property P we are testing for then S_x will always contain a number of elements called 'witnesses'. A witness is an entity whose existence shows that x does not in fact possess the property P. For instance, a factor $x > a > 1$ of an integer x is a witness to x's not having the property of being prime. Each test then works by randomly selecting an element $a \in S_x$ and checking whether a is a witness. If a is not a witness, then x passes the test for that iteration of the algorithm. If it is a witness, the algorithm outputs 'no'.[134] If witnesses are known to be abundant for elements that lack the property P, the fact that an object x has passed many such tests provided good evidence that x does have the property P.

Monte Carlo algorithms for decision problems are subject either to two-sided errors, wherein the program may be in error when answering either 'yes' or 'no', or to one-sided errors, wherein only receipt of one of the two outputs is inconclusive. However, the kind of IMC algorithm just described is only subject to one-sided errors: the output 'no' is always conclusive, but the output 'yes' is not so. This is because not every element of S_x need be a witness, so we may incorrectly come to believe that x has the property P by being unlucky and missing all the witnesses when we randomly select elements from S_x. However, for the cases we are interested in we will be able to attach a quantitative value to the probability of this happening. In the next section, we continue exploring IMC algorithms by considering a class of methods for verifying a wide variety of mathematical claims: the hypergeometric algorithmic identity theory of the mathematician Doron Zeilberger.

[133] Ibid., 64.
[134] Richard M. Karp, "An Introduction to Randomized Algorithms".

3.ii. Algorithmic Identity Theory and Polynomial Comparison

Doron Zeilberger has discovered a very general method for proving identities between sums of generalised hypergeometric q-hypergeometric terms.[135] In the most basic kind of case, a function $F(n,k)$ is hypergeometric if $\frac{F(n+1,k)}{F(n,k)}$ and $\frac{F(n,k+1)}{F(n,k)}$ are each rational functions of n and k, and is a q-hypergeometric term if these ratios are rational functions of q, q^k and q^n.

An example of a hypergeometric identity is:

Identity 3.2.

$$\sum_{k=-n}^{n}(-1)^k\binom{2k}{n+k}^3=\frac{(3n)!}{n!^3}$$

Two examples of q-hypergeometric identities are as follows:

Identity 3.3.

Let $(q)_r := (1-q)(1-q^2)\ldots(1-q^r)$. Then

$$\sum_{r=0}^{n}\frac{q^{r^2}}{(q)_r(q)_{n-r}}=\sum_{r=-n}^{n}\frac{(-1)^r q^{(5r^2-r)/2}}{(q)_{n-r}(q)_{n+r}}$$

Identity 3.4.

Let H_n be given by $H_n = H_n(q) = \frac{(1+q)}{(1-q)}\frac{(1+q^2)}{(1-q^2)}\ldots\frac{(1+q^n)}{(1-q^n)}$. Then

$$\left(\sum_{k=0}^{n}\frac{2(-q^{n+1})^k}{1+q^k}H_k\right)^4\sum_{k=-n}^{n}\frac{4(-q)^k}{(1+q^k)^2}\frac{H_{n+k}}{H_n}\frac{H_{n-k}}{H_n}=\left(\sum_{k=-n}^{n}(-q)^{k^2}\right)^4$$

Letting n tend to infinity, Zeilberger derives the following results:

Identity 3.5.

Let $(q)_r$ be as in 7. Then

$$\sum_{r=0}^{\infty}\frac{q^{r^2}}{(q)_r}=\prod_{i=0}^{\infty}\left(1-q^{5i+1}\right)^{-1}(1-q^{5i+4})^{-1}$$

[135] Doron Zeilberger, "Theorems for a Price", *Notices of the AMS* 40 (1993): 978-981.

Identity 3.6.

$$\left(\sum_{k=-\infty}^{\infty} q^{k^2}\right)^4 = 1 + 8\sum_{k=1}^{\infty} \frac{q^k}{(1+(-q)^k)^2}.$$

These last two identities are themselves equivalent to two famous theorems of number theory. Identity 3.5 is equivalent to the first Rogers-Ramanujan identity, which asserts that the number of partitions of an integer into parts that leave remainder 1 or 4 mod 5 is equal to the number of partitions into parts that differ from each other by at least 2. Identity 3.6 is equivalent to a theorem of Jacobi that says that the number of decompositions of an integer into the sum of 4 squares is equal to 8 times the sum of all its divisors other than those that are multiples of 4.

Hypergeometric and q-hypergeometrics identities encompass or are equivalent to a huge range of mathematical theorems. Later Zeilberger showed that his techniques can be applied to sums involving more variables, and to integrals as well as sums, extending their scope even further.[136] The approach can now be used on 'most of the identities between the classical special functions of mathematical physics', and indeed 'all "natural identities" we are now aware of'.[137]

Although we shall not go into too much detail here, essentially the idea is that the algorithm gives a way of reducing the proof of any such identity to that of proving a finite identity amongst rational functions, and hence to an identity between specific finite polynomials by multiplying up denominators. Zeilberger then further reduces this to showing the existence of a solution to a large system of inhomogeneous linear equations with symbolic coefficients – i.e., with coefficients that are themselves functions of further auxiliary variables.

Systems of linear equations with numerical coefficients are easy to solve, but if the coefficients are symbolic then the problem can swiftly become computationally intractable. Zeilberger therefore suggests proceeding as follows.[138] We substitute numerical values for the variables comprising the symbolic coefficients, and then check whether the corresponding system of numerical equations is soluble. If it is not, we have found witnesses to the identity being false. However, suppose that we continue to substitute in further values for the symbolic coefficients, and this yields a soluble system each time. This strongly suggests that the symbolic system also has a solution, as it is unlikely that an arbitrary system of inhomogeneous linear equations with more equations than variables will be soluble.

Next we give a related but much simpler example of an IMC algorithm, where in this case we can associate the procedure with a determinate probability of error. Suppose we have two complicated algebraic expressions for a pair of real polynomials $g_1, g_2 \in \mathbb{R}[x]$, each of degree d. Suppose further that we believe,

[136] Doron Zeilberger and Herb Wilf, "An Algorithmic Proof Theory for Hypergeometric (Ordinary and 'q') Multisum/Integral Identities", *Inventiones Mathematicae* 108 (1992): 575-633.
[137] Zeilberger, "Theorems For a Price"
[138] Ibid.

though perhaps are not certain, that $g_1 \equiv g_2$. We have the following algorithm for testing this claim, discovered independently by Jack Schwartz and Richard Zippel.[139]

Firstly, define $f = (g_1 - g_2)$. Clearly, our claim is equivalent to the assertion that f is the zero polynomial. If there is some $r \in \mathbb{R}$ such that $f(r) \neq 0$, then r is a witness to the falsity of this conjecture. Moreover, if f is non-zero then it has degree at most d, and hence at most d roots in $\mathbb{R}$, with every other element of $\mathbb{R}$ being a potential witness. Now let $S = \{\pm 1, \pm 2, \dots, \pm d\}$. If f is not the zero polynomial, then at least half of the members of S will be witnesses. We therefore pick a finite sequence $(r_1, r_2, \dots, r_n)$ of n members of S independently and at random, and check each of them to see whether they are witnesses. Assuming that f is not the zero polynomial, and that the r_i are selected independently, we have the following result:

Lemma 3.7. (Schwartz-Zippel)

Let f be a non-zero polynomial of degree at most d and $\{r_1, \dots, r_n\}$ be selected randomly and independently from $S = \{\pm 1, \pm 2, \dots, \pm d\}$. Then:

$$\mathbb{P}(f(r_1) = 0, f(r_2) = 0, \dots, f(r_n) = 0) \leq \frac{1}{2^n}$$

Clearly, as we increase *n* it becomes increasingly unlikely that any given non-zero polynomial f will pass all *n* tests. So if f passes every test for some sufficiently long finite sequence $(r_1, r_2, \dots, r_n)$ we have randomly generated, then this strongly suggests that the condition on f used to derive the above probability statement must be false: that is, that f must be the zero polynomial.

3.iii. Hypothesis Testing and Statistical Inference

Let us now see the Schwartz-Zippel algorithm in action in the context of a concrete example. Suppose that g_1 is defined as follows:

$$g_1(t) = (x^2 + 13x)^6 - (2x + 3)^5 - (4x - 1)^2 + 13$$

Suppose further that we want to know its coefficients explicitly. I attempt to expand the brackets by hand and get the following polynomial as a result:

$$g_2(t) = x^{12} + 78x^{11} + 2535x^{10} + 43940x^9 + 428415x^8 + 2227758x^7 + 4826809x^6 - 32x^5 - 240x^4 - 720x^3 - 1096x^2 - 802x - 231$$

If I have done the expansion correctly, then we should have that $g_1 \equiv g_2$, so that f will be the zero polynomial and no witnesses to the claim $f \not\equiv 0$ will exist. Yet so much numerical calculation and symbolic manipulation was involved in the

[139] Zippel, "Probabilistic Algorithms for Sparse Polynomials". Jack Schwartz, "Fast Probabilistic Algorithms for Verification of Polynomial Identities", *Journal of the ACM* 27 (1980): 701-717.

algebra that the possibility of a mistake having been made when expanding the brackets cannot be ruled out. This given, it is clearly useful to have a way of checking that the expansion is correct. Now, let $S = \{\pm 1, \pm 2, \ldots, \pm 12\}$. We generate 15 members of S at random using a random number generator and calculate the value of $f(r_i)$ in each case.

r_i	$g_1(r_i)$	$g_2(r_i)$	$f(r_i)$
4	98867321361	98867321361	0
-3	729000087	729000087	0
2	728983157	728983157	0
12	728999985648897	728999985648897	0
9	60254725476351	60254725476351	0
1	7526415	7526415	0
-10	730418189	730418189	0
3	12230531307	12230531307	0
7	7529534579427	7529534579427	0
-10	730418189	730418189	0
-10	730418189	730418189	0
-6	5489090181	5489090181	0
-3	729000087	729000087	0
-12	7067697	7067697	0
-11	115853991	115853991	0

Intuitively speaking, the fact that no witnesses were found after fifteen attempts is highly persuasive unto itself: given this evidence, it seems very unlikely – though not strictly impossible – that f is not the zero polynomial. But further to this we can also use Lemma 3.7 to attach a quantitative value to the evidence. If f were a non-zero polynomial, the probability of it having passed all of our tests would be less than $\frac{1}{2^{15}} \cong 0.0000305$. This would be a somewhat miraculous occurrence, so it is reasonable to believe that I have expanded the brackets correctly after all.

This argument does seem convincing. But as with the quasi-empirical evidence for the Goldbach Conjecture in the previous chapter, we should spell out more carefully what the underlying rules of inference consists in, particularly with regards to the significance of the value we have assigned to the probability of a false positive. Happily, this time we do have an established inferential framework to appeal to. The structure of the argument is essentially that of the hypothesis testing or 'test of significance' procedure pioneered by the great statistician Sir Ronald Fisher. The method runs as follows.[140]

[140] This is the method for a one-tailed test where if the null hypothesis H_0 is false then t may be expected to take a larger value than if it were true. In some cases a two-tailed test also considering $\mathbb{P}(t \leq t_{obs}|H_0)$ is preferable, though this need not concern us here.

Statistical Hypothesis Testing

1. Decide upon a level of significance, α.

2. State the null hypothesis H_0, and an alternative hypothesis H_1.

3. Decide upon a test statistic t, the observed value of which will determine whether to accept or reject the null hypothesis H_0.

4. Give a sampling procedure that will yield an observed value of t.

5. Determine the distribution of t under the sampling procedure, assuming that H_0 is true.

6. Carry out the sampling procedure to give an observed value, t_{obs}.

7. If $\mathbb{P}(t \geq t_{obs}|H_0) < \alpha$ then reject the null hypothesis H_0.

The level of significance α here denotes the probability that H_0 is rejected, given that it is in fact true. This gives us the following procedure for the polynomial comparison task. Let $g_1, g_2 \in \mathbb{R}[t]$ each be of degree d as before. In the description given below the test statistic t can take on only two values, namely 0 or 1. Its distribution is given by Lemma 3.7.

Algorithm 3.8. (Modified from the Schwartz-Zippel algorithm)

1. Select $0 < \alpha \ll 1$, the desired level of significance. Let H_0 be the hypothesis that $g_1 \neq g_2$ and H_1 that $g_1 = g_2$. Define $n = \lceil -\log_2 \alpha \rceil$.

2. Pick an n-tuple $(r_1, r_2, \dots, r_n)$ at random from $S = \{\pm 1, \pm 2, \dots, \pm d\}$

3. Compute $s = \sum_{i=1}^{n} (g_1(r_i) - g_2(r_i))^2$. Let $t = 1$ if $s = 0$ and $t = 0$ otherwise.

4. If $t = 0$, then H_0 is true and $g_1 \neq g_2$. If $t = 1$ then H_0 is rejected at the level of significance α.

Let us examine the extent to which this procedure meets the demands articulated in Section 2.vi for an objective, determinate way of assessing the quality of our evidence. Though α must be chosen in advance, a potential sceptic is free to choose any value they like. Scientists commonly work with the values $\alpha = 0.05$, or perhaps $\alpha = 0.01$. But even the rather modest investigation I carried out for the bracket expansion task could have accommodated an α that was 300 times smaller than this latter value. Computers are so fast at performing these algorithms that millions of tests can be run in a short space of time, so that given any algorithm that has the iterative structure above, leading to an exponential decrease in

$\mathbb{P}(t \geq t_{obs}|H_0)$, and if H_0 actually is false, we can always acquire sufficient evidence for the rejection of H_0 for any reasonable choice of α.

An issue arises, however, in moving to the conclusion that the alternative hypothesis H_1 is true, given that H_0 has been rejected. Whether we are justified in doing this depends on further assumptions about the testing framework: most crucially, it is sensitive to the prior plausibility of H_1 in comparison to H_0. For it might be that $H_1$is *prima facie* so unlikely that acceptance of it becomes distinctly unattractive even if H_0 has been rejected at a highly demanding level of significance. We may compare this with the mathematics of clinical trials that scan for diseases. Even if reliable tests with only a small chance of giving false positives are used, in cases of a condition that is extremely rare it may still be unlikely that a person with positive results actually has the disease in question.

This given, hypothesis testing on its own will not enable us to produce a self-contained argument for the conclusion $g_1 = g_2$. The experiment performed above did seem intuitively convincing, however, so let us see how the argument might be patched up. What is required is a Bayesian approach whereby we make a quantitative assessment of the prior likelihood of H_0 being true, which is then updated in light of the new evidence furnished by the hypothesis testing procedure. But although this technique is of immense value in the empirical sciences, I shall argue that within mathematics it is unsuitable in the context of Public Acceptance.

One reason for this unsuitability is of a technical nature. The derivation of 'almost all the important results'[141] of Bayesian statistics requires an assumption of logical omniscience. But because the identity of two finite polynomials is an *a priori* truth, it follows that we should know the answer in advance. It may be possible to give a humanist version of Bayesianism where we can attach non-trivial probabilities to mathematical truths. However, this is not the approach we will take in this thesis, because there are at least two other problems as well.

Most importantly, there is the notorious problem of determining values for the prior probabilities of the null and alternative hypotheses. Doing this correctly here requires us to know how the candidate polynomials have been arrived at, which is often inscrutable. We might think it would involve an assessment of my personal reliability in multiplying out brackets of this complexity. But perhaps I didn't multiply them out by hand at all, and simply used a computer to do the expansion for me! Speaking more generally, because such assessments are always merely subjective, and never determined by a definite procedure, disagreement will inevitably arise about their allocation: this will then be damaging to Consensus.

Secondly, admitting propositions expressing epistemic probability statements into the body of Publicly Accepted mathematics is also problematic because they are not stable over time. After we have made an assessment of the prior probability, asserting a proposition such as $\mathbb{P}(H_0) = 0.6$, we update this assessment in light of the evidence produced by the testing phase and give a statement of the posterior

[141] William Talbott, "Bayesian Epistemology", *Stanford Encyclopedia of Philosophy*, 6.1A, accessed 11th August 2015, http://plato.stanford.edu/entries/epistemology-bayesian

probability, such as $\mathbb{P}(H_0) = 0.03$. Again, we may then do further testing, causing us to update this statement and give the contradictory proposition $\mathbb{P}(H_0) = 0.002$.

For the kinds of problems we will now restrict our attention to, however, we will soon see that an approach based on ideas from classical or 'frequency' statistics is available. This avoids these problems without introducing any further apparatus than we have already met so far: the techniques given below will require only the axioms of probability and the concept of random selection from a finite set.

For the remainder of the chapter, we will focus on a version of the Rabin-Miller Algorithm. This will give us a Monte Carlo algorithm for finding a prime number in a certain range – as opposed to an IMC algorithm that checks if a certain fixed number n is prime, which for reasons given above would require us to give a 'prior probability' of n's being prime. This algorithm will work by randomly selecting a candidate for primality from a predetermined class, and only then subjecting this candidate to IMC testing. It will continue to select further candidates until one is found which passes all of the IMC tests: it is thus an iterated iterated Monte Carlo algorithm.

Because we will give lower bounds for the number of prime numbers in the class from which candidates are drawn, we will also be able to give a lower bound for the probability that we select a prime number as a candidate. The procedure can therefore be associated with a definite, comprehensive probability of error, representing the chances that the algorithm outputs a composite rather than prime number, without us ever having to assign a prior probability to a specific number's being prime. Moreover, will also see that the probability of error can be made smaller than any given positive number by picking an appropriate number of tests.

3.iv. The Rabin-Miller Algorithm

In recent decades, there has been tremendous interest in finding efficient algorithms for discovering large prime numbers. Records are kept of the current largest known primes, all of which have been found by The Great Internet Mersenne Prime Search. This project, which has been running since 1997, is a collaborative effort whereby around 100,000 volunteers have downloaded free software onto their personal computers so as to contribute to the available computing power.[142] This has led to the discovery of all of the recent largest known primes, including the current largest $2^{57,885,161} - 1$, which was discovered on 25th January 2013.[143]

Of course, being considered part of mathematics, these achievements are subject to the usual restriction to deductive methods discussed in Chapter 1. However, finding a steady supply of large prime numbers is also of great practical

[142] "Mersenne Number", *Encyclopedia Britannica Online*, accessed 11th August 2015, http://www.britannica.com/topic/Mersenne-number

[143] "GIMPS Discovers 48th Mersenne Prime, $2^{57885161} - 1$ is Now the Largest Known Prime", *Great Internet Mersenne Prime Search*, accessed 11th August 2015, http://www.mersenne.org/primes/?press=M57885161

importance in modern technological industry – not least because it is necessary for maintaining the security of public-key encryption, a method of coding information that underlies the data transactions upon which our financial institutions and infrastructure crucially depend. This given, there has also been a great deal of interest in the discovery of 'industrial grade' primes, a phrase coined by H. Cohen.[144] This is not a determinate mathematical property of a number, but rather indicates that it has been selected by a procedure that is sufficiently likely to output prime numbers for practical purposes.

In 1977, Robert Solovay and Volker Strassen developed an IMC algorithm for testing whether a given number was prime.[145] This was of great importance as at the time there were no efficient deterministic algorithms available for this task. Similarly to the Schwartz-Zippel algorithm given above, it is such that the probability that a composite number n passes all the tests can be precisely bounded, and in principle made as small as desired by running a sufficient number of tests. Michael O. Rabin later responded by showing that a deterministic primality test[146] provided by Miller in 1976 – which depended on the generalised Riemann Hypothesis being true – could be adapted to produce another, more efficient algorithm for the same task that worked unconditionally, and also within precisely specifiable error bounds.[147]

In his 1980 paper entitled 'Probabilistic Algorithm for Testing Primality', Rabin describes the algorithm and reports some experiments carried out with Vaughan Pratt.[148] We are told that at the time 'the computations were done on a medium-sized computer' and that 'numbers with several hundred binary digits were generated and tested', all taking at most a few minutes to run.[149] Rabin used his algorithm to correctly identify which of the Mersenne numbers $2^p - 1$ were also Mersenne primes for prime $p < 500$. He also correctly identified $2^{300} - 153$ as the largest prime less than 2^{300}, and likewise $2^{400} - 593$ as the largest prime less than 2^{400}, as well as discovering what were at the time the largest known pairs of twin primes: $\left(\prod_{p_i<300} p_i\right)\times 338 + 821$ and $\left(\prod_{p_i<300} p_i\right)\times 338 + 823$.

Let us now examine the Rabin-Miller Algorithm in depth. For any integer n, we define $S_n = \{1, \dots, n\}$. We first prove the following lemma:

Lemma 3.9.

Let $p \geq 3$ be prime, and let $p - 1 = 2^s r$, where r is odd and $s \geq 1$, and let $a \in S_{p-1}$.

[144] Crandall and Pomerance, *Prime Numbers*, 126.
[145] Robert Solovay and Volker Strassen, "A Fast Monte-Carlo Test for Primality", *SIAM Journal on Computing* 6 (1977): 84-85.
[146] Gary Miller, "Riemann's Hypothesis and Tests for Primality", *Journal of Computer and System Sciences* 13 (1976): 300-317
[147] Michael Rabin, "Probabilistic Algorithm for Primality Testing", *Journal of Number Theory* 12 (1980): 128-138
[148] Ibid.
[149] Ibid., 136.

Then either:

1.) $a^r \equiv 1 \mod p$, or

2.) $\exists\, i \in \{0, \ldots, s-1\}$ such that $a^{2^i r} \equiv -1 \mod p$.

To prove this, we will need two preliminary results. Firstly, recall Fermat's Little Theorem (FLT) from Chapter 1: if p is prime then $a^p \equiv a \mod p$. Moreover, if $(a, p) = 1$ as is the case here then we can multiply by the inverse of a (modulo p) to get $a^{p-1} \equiv 1 \mod p$.

Next, suppose p is prime and that $x^2 \equiv 1 \mod p$. Then

$$(x-1)(x+1) \equiv 0 \mod p$$

$$\Rightarrow \quad p \mid x-1 \ \text{or} \ p \mid x+1$$

$$\Rightarrow \quad x \equiv 1 \ \text{or} \ -1 \mod p.$$

That is, ± 1 are the only square roots of 1, modulo a prime p. Now, returning to Lemma 3.5, we know that

$$a^{2^s r} \equiv 1 \mod p$$

from the FLT. We then need only take successive square roots, modulo p. Either we get the value -1 on the right hand side at some stage in this process, in which case the second condition holds, or if not we are left with a +1 at the end of the process, after taking square roots s times, in which case the first condition holds. In either case, one of the conditions given in Lemma 3.9 will be met, as required.

The converse of Lemma 3.9 does not hold, however. Given a composite number $n = 2^s r + 1$, there may be integers $a \in S_{n-1}$ such that neither $a^r \equiv 1 \mod n$ holds, nor is there an $i \in \{0, \ldots, s-1\}$ with $a^{2^i r} \equiv -1 \mod n$. Here we say that n is a 'pseudoprime to the base a', and that a is a 'strong liar' with respect to n.

In keeping with the usage introduced in Section 3.i., we call an $a \in S_n$ a 'witness' for a composite number n when a is not a strong liar with respect to n: that is, when it meets neither of the two conditions given in Lemma 3.9. The existence of a witness therefore shows us that n is indeed composite. Moreover, in his 1980 paper, Rabin showed that if $n \geq 5$ is composite number then at least ¾ of the positive integers less than n are witnesses to n being composite. In fact, in most cases the number of witnesses tends to be far higher, but this suffices as a global upper bound. The proof is several pages long and so will be omitted here, but interested readers are referred to his paper for details.[150]

[150] Rabin, "Probabilistic Algorithms for Primality Testing", 130-133.

Give an integer $n \geq 5$, we therefore have the following IMC algorithm for testing whether n is prime. We generate a succession of positive integers $(a_1, a_2, \ldots, a_k)$ from S_{n-1}. The integer n passes the i^{th} iteration of the test if a_i is not a witness to the compositeness of n. As suggested above, the algorithm asserts n to be composite if it fails any of the k tests, and if it passes all of them then it asserts n to be prime. As the tests are independent, Rabin's result gives that the probability that a composite integer n is falsely asserted to be prime by this IMC algorithm is less than $\frac{1}{4^k}$. As before, we can make this smaller than any given positive probability α, this time by setting $k = \lceil -\log_4 \alpha \rceil$.

Now, to construct our Monte Carlo algorithm for finding a prime number from a set P, which must be known to contain at least a certain proportion of primes, we generate a sequence $\langle n_i \rangle$ of candidates for primality by selecting them randomly and independently from P. We subject each successive candidate n_i to k iterations of the IMC test just described, and the first candidate to pass all tests is asserted to be prime by the algorithm. As we shall see below, we will be able to assign a definite upper bound to the probability that this algorithm fails to output a prime number. And as promised above, the procedure never makes use of claims of the form $\mathbb{P}(n \; is \; prime) = \alpha$, where n is a specific number and α is a non-trivial probability. Indeed, Rabin himself expresses the view that such claims are 'nonsensical since n is either prime or not.'[151]

3.v. Rabin-Miller in Action

Consider the following experiment. Suppose that we have been informed by an oracle – that we may assume to be infallible – that precisely two of the five numbers in the set $S = \{503,601,703,803,901\}$ are prime. Suppose further that we decide to pick numbers at random from S, and subject them to 5 iterations of the Rabin-Miller Algorithm until we have a number – which could be prime or composite – that has passed all 5 tests. The program could in theory run forever: we could keep getting composite numbers that fail at least one of the tests. However, it clearly terminates with probability 1.

When we run the algorithm, there are two things that can happen. When the compiler selects candidates for testing, it will hopefully be the case that the first one to pass all the tests is prime, so that the algorithm gives a true prime number as output. Alternatively, we could be unlucky: the compiler might pick only composite numbers until one such candidate passes all five tests, because at the IMC testing phase only strong liars are selected for this candidate. In this case, the program will incorrectly assert this candidate to be prime.

Now, the probability of picking a composite candidate is always $3/5$ each time, and Rabin's bound gives us that the chances of any composite number passing the test 5 times is at most $\frac{1}{4^5} = \frac{1}{1024}$. Thus, the chances of picking n composite numbers in a row that each fail the test before finding a prime is therefore at least:

[151] Ibid., 129.

$$\left(\frac{3}{5}\right)^n \left(\frac{1023}{1024}\right)^n \frac{2}{5}$$

By summing over n and using the formula for geometric series we have the following result:

$$\mathbb{P}(Output\ is\ prime) \geq \frac{2}{5}\sum_{n=0}^{\infty}\left(\frac{3069}{5120}\right)^n = \frac{2}{5}\ \frac{1}{1-\frac{3069}{5120}} = \frac{2048}{2051}$$

and hence that

$$\mathbb{P}(Output\ is\ composite) \leq \frac{3}{2051}$$

Suppose now that I run the algorithm and the output is 601. Clearly, we can now be fairly confident that this number is prime. However, if I were an engineer that used this kind of reasoning as a belief-formation mechanism 2051 times in a year, I could expect to make around 3 mistakes. If each of these corresponded to one of my bridges collapsing, our procedure would not be sensitive enough for this application.

Next, let us generalise this procedure; suppose we have q integers, of which at least p are known to be prime ($1 \leq p < q$). As before, we pick candidates at random from this set, running the Rabin-Miller test k times on each, and outputting the first number to pass all the tests. The probability of getting n composite numbers that are rejected by the test before finding a prime is therefore at least

$$\left(\frac{q-p}{q}\right)^n \left(\frac{4^k-1}{4^k}\right)^n \frac{p}{q}$$

Summing as before, we have that

$$\mathbb{P}(Output\ is\ prime) \geq \sum_{n=0}^{\infty}\frac{p}{q}\left(\frac{(q-p)(4^k-1)}{4^k q}\right)^n$$

$$= \frac{p}{q}\ \frac{1}{1-\frac{(q-p)(4^k-1)}{4^k q}} = \frac{p4^k}{p4^k+q-p}$$

So that

$$\mathbb{P}(Output\ Composite) \leq \frac{q-p}{p4^k+q-p} < \frac{q/p}{4^k}$$

for $p \geq 1$. Hence as long as we know that there is at least one prime in our set we can make the chances of error in our algorithm as small as we like by increasing k. More precisely, we have the following result:

Theorem 3.10.

Let ε belong to the open interval $(0,1)$, and let S be a finite set of q integers, at least p of which are prime. Define $k = \left\lceil \log_4 \left(\frac{q}{\varepsilon p}\right) \right\rceil$. If we run the Rabin-Miller Algorithm with k iterations on successive randomly selected elements of S and output the first to pass all k tests, then we have that $\mathbb{P}(Output\ Composite) < \varepsilon$.

When we are looking for numbers that are likely to be prime, the situation is of course unlike the artificial example above in that we do not have an infallible oracle to tell us that a certain number of our potential candidates are prime. However, we can resolve this issue by using one of the many explicit bounds on the number of primes $\leq x$, which is denoted by $\pi(x)$. For example, when $x \geq 55$ the following inequality holds:[152]

$$\frac{x}{\ln(x)+2} < \pi(x) < \frac{x}{\ln(x)-4}$$

Suppose that $S = \{2^n, 2^n+1, \ldots, 2^{n+1}-1\}$, the set of numbers with $n+1$ binary digits. Then we have that $q = 2^n$. We obtain a value for p, the expression giving a lower bound for the set of primes in S, as follows. The actual the number of primes in S is equal to $\pi(2^{n+1}-1) - \pi(2^n - 1)$. Using Rosser's result, taking the lower bound for the first term and the upper bound for the second, we have that the number of prime elements in S is at least

$$\frac{2^{n+1}-1}{\ln(2^{n+1}-1)+2} - \frac{2^n-1}{\ln(2^n-1)-4}$$

We can therefore replace q/p in the above expression with the larger expression

$$\left(\frac{2-1/2^n}{\ln(2^{n+1}-1)+2} - \frac{1-1/2^n}{\ln(2^n-1)-4}\right)^{-1}$$

We have proved the following theorem:

[152] Barkley Rosser, "Explicit Bounds for some functions of prime numbers", *American Journal of Mathematics*, 63 (1941), 211.

Theorem 3.11.

Let $n \geq 3$, let $S = \{2^n, 2^n + 1, \ldots, 2^{n+1} - 1\}$, and let ε belong to the open interval $(0,1)$. Define the function $f: \mathbb{Z}^+ \to \mathbb{R}^+$ by the formula

$$f(n) = \frac{2 - 1/2^n}{\ln(2^{n+1} - 1) + 2} - \frac{1 - 1/2^n}{\ln(2^n - 1) - 4}$$

Define $k = \left\lceil \log_4 \left(\frac{1}{\varepsilon f(n)} \right) \right\rceil$. We select successive random numbers from S and subject them to k iterations of the Rabin-Miller Algorithm, outputting the first number to pass all k tests.

Then we have that $\mathbb{P}(Output\ Composite) < \varepsilon$.

Lastly, we illustrate Theorem 3.11 with an example. Suppose that I want to find a prime number that is at least 2^{100} but less than 2^{101}. Suppose further that I am very demanding and will tolerate a probability of error of at most 10^{-100}. We substitute the values $n = 100$ and $\varepsilon = 10^{-100}$ into the above formula, giving $f(n) \cong 0.012464$ and $k = 170$. So running 170 iterations of the IMC test for each candidate will yield a procedure meeting the desired level of accuracy.[153]

3.vi. *Intuition and Cognitive Bias*

In this section, we use both a deductive algorithm and the procedure just described to try to find prime numbers p satisfying $2^{25} \leq p < 2^{26}$. The first algorithm – which is based on the Trial Division Algorithm, and always gives a prime number when performed correctly – is a Las Vegas procedure that runs as follows.

Algorithm 3.12.

1. Randomly pick a number n with $2^{25} \leq n < 2^{26}$
2. Compute the residue of n modulo k for all integers k with $2 \leq k \leq \lfloor \sqrt{n} \rfloor$
3. If $n \not\equiv 0 \bmod k$ for all k with $2 \leq k \leq \lfloor \sqrt{n} \rfloor$, assert n to be prime. If $n \equiv 0 \bmod k$ for some such k, return to step 1.

We now consider a Monte Carlo algorithm for the same task. Putting $n = 25$ and $\varepsilon = 10^{-10,000}$ into Theorem 3.13 we get the value $k = 16{,}613$. Hence, the following Monte Carlo algorithm produces composite output with probability less than $10^{-10,000}$.

[153] For some interesting empirical results connecting the probability of error with the number of tests, see Ivan Damgård, Peter Landrock and Carl Pomerance, "Average Case Error Estimates for the Strong Probably Prime Test", *Mathematics of Computation* 61 (1993): 177-194.

Algorithm 3.13.

1. Randomly pick a number n with $2^{25} \leq n < 2^{26}$

2. Subject n to 16,613 iterations of the Rabin-Miller test.

3. If n passes all 16,613 tests, assert n to be prime. Otherwise return to step 1.

Suppose now for some important practical purpose we need to acquire a prime number – our lives are threatened by a demon, say, who will kill us unless we supply him with a prime in this range. Suppose also that we have a physical computer programmed with both algorithms, and so can use either of them to supply what we need. Which one should we trust? As an intuition pump,[154] I have programmed both algorithms in C on my laptop. After extensive testing, wherein I convinced myself there were no errors remaining in the code, the algorithms were run once each, and the following results produced:

- Algorithm 3.12 outputted the number $\boldsymbol{n = 48,140,819}$

- Algorithm 3.13 outputted the number $\boldsymbol{n = 66,998,713.}$

The reader is now requested to choose one of the two numbers, imagining that picking a prime is necessary in order to appease the daemon, and to record their choice by writing it down on a piece of paper. My intuitive reactions were as follows. With respect to the first integer, I now feel completely sure that it is a prime number: checking it against an online database would add nothing to my confidence, for example. With regards to the second number, however, I still feel a lingering sense of anxiety, exactly analogous to that induced by the quasi-empirical evidence in the previous chapter. Were I to now check and find that this number is indeed prime, I would likely feel a sense of relief, as would someone who had taken a risk that had proved to pay off: a sense of reassurance exactly analogous to the feeling met with in Chapter 2 when deductive proofs were supplied to corroborate the non-deductive evidence.

These intuitions fit well with mathematicians' actual attitudes towards the two algorithms with regards to Public Acceptance. Being Monte Carlo, Algorithm 3.13 would never suffice to justify a claim in a peer-reviewed journal, but Algorithm 3.12 would be acceptable in this regard. Yet from a normative epistemological point of view, these intuitions are surely bizarre! Rabin remarks in his paper that even his much larger error value of 10^{-18} 'seems small when compared to the frequency of machine errors present in practical computations'.[155] There are also many other ways in which the numbers produced might turn out to be composite:

[154] In Dan Dennett's positive sense of a thought experiment designed to focus our intuitive responses on the important features of a problem. See for example Daniel Dennett, *Intuition Pumps and Other Tools for Thinking* (New York: W. W. Norton and Company, 2013).
[155] Rabin, *"Probabilistic Algorithm for Primality Testing*, 135.

perhaps I have made a programming error or transcribed a number incorrectly from the compiler output to this manuscript.

If $66{,}998{,}713$ does turn out to be composite, and the hardware is somehow known not to have malfunctioned, then at this point even the most enthusiastic reader must find it more likely that I have fabricated the evidence rather than that the asserted sequence of events – the number passing 16,613 tests – had actually occurred, especially if they have been exposed to Hume's arguments in 'On Miracles'. Indeed, even our eyesight will fail us with a probability far higher than $\varepsilon = 10^{-10,000}$, a number that is so small as to defy practical comprehension. Put simply, the Monte Carlo nature of Algorithm 3.13 should be the least of our worries with regards to error, rather than being the sole determinant of such contrasting intuitive responses.

We turn now to giving a partial psychological explanation of that fact that I – and perhaps the reader, too – felt far more comfortable with Privately Accepting that $48{,}140{,}819$ was prime rather than that $66{,}998{,}713$ was prime, and would have felt far more comfortable giving this response to the daemon. Philosophers often treat such intuitive reactions as an important source of evidence, and in many cases this may be reasonable. However, the number $\varepsilon = 10^{-10,000}$ is so small that our intuitive judgements are unlikely to be reliable. Quantities of this scale are well outside of our experience, and so we should not expect to have developed a sufficiently fine-grained intuitive sensitivity to the actual risk of error here.

Let us consider the magnitude of this number a little further. We attempt to put the quantity ε into perspective by describing three events that would each occur with roughly this probability:

- I take a job at a casino. I work there 5 days a week for 24 years. Every working day, as I start my shift, I place money on the roulette wheel landing on 13 Black. I win every time I play.

- I decide to spend my retirement playing blackjack with friends. This involves me shuffling decks of cards frequently. I do this five times a day for a month. I give the cards a thorough shuffle. However, after every single shuffling, the cards are always left partitioned into suits, with the cards in each suit perfectly arranged in ascending order.

- I try playing the UK national lottery. I buy one ticket each Saturday. I continue this every week for 27 years. I win the jackpot every time.

These examples might now give the reader a fuller sense of just how unlikely it is that we have missed all the witnesses when running Algorithm 3.13, and hence that a computer error in either algorithm would be a far more likely occurrence: a conclusion that will be argued for carefully in the next chapter. Yet even these more tangible scenarios are still quite hard to understand intuitively, and one of

these three events may seem more or less likely than the other two, though in fact they are roughly equally likely to occur.[156]

In what must be one of the most widely cited papers in the social sciences, Daniel Kahneman and Amos Tversky discuss heuristic mechanisms that humans use to arrive at judgements when the complexity of a problem makes explicit calculation impossible.[157] They found it quite easy to construct situations where the answers given by these heuristics deviates from normative statistical theory, and their results have been replicated in many experiments since. It is therefore clear that human judgements are vulnerable to systematic error or cognitive biases – just as the intuitive responses induced above make little sense from normative epistemology considerations. In particular, these researchers found that humans are especially bad intuitive statisticians, because our experience is not coded in a way that allows us to learn to correct our own errors here (indeed, even *statisticians* were later found to be bad intuitive statisticians![158]).

We leave the detailed explanation of faulty judgements of reliability in this particular case as an open question, noting only that cognitive psychology both suggests it is not surprising and appears to possess the resources needed for an explanation. I will however tentatively mention one line of thought that might be relevant. One means of coming to intuitive judgements that is discussed in the aforementioned paper is the resemblance heuristic, whereby we compare the profile of a given case about which a judgement must be made to what we take to be properties of a general class. In a famous experiment, they present the following question to a number of subjects:

> 'An individual has been described by a neighbor as follows: "Steve is very shy and withdrawn, invariably helpful but with little interest in people or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail." Is Steve more likely to be a librarian or a farmer?'[159]

Most subjects answered that the person was more likely to be a librarian, because the neighbour's description more closely resembles a stereotypical description of someone drawn from this profession. Yet there are actually more than 20 times as many male farmers as librarians in American, where the study was conducted, and as we saw earlier in the chapter this base rate must be taken into account. In this case, this makes 'farmer' a more reasonable answer. A relevant factor is thus ignored, and so the resemblance heuristic can lead us astray in some cases.

In the above experiment, the number 48,140,819 was produced by a deductive algorithm, whereas 66,998,713 was produced by an algorithm whose working is known not be deductively valid. Moreover, we have already encountered many

---

[156] This suggestion was confirmed through an informal experiment wherein I asked some friends to express an opinion about the relative likelihoods. Though there was no clear consensus, many respondents felt one of the events was much less likely than the others.

[157] Daniel Kahneman and Amos Tversky, "Judgement under Uncertainty: Heuristics and Biases", *Science* 185 (1974): 1124-1131.

[158] Daniel Kahneman, *Thinking Fast and Slow*, (London: Penguin, 2011), 5.

[159] Kahneman, *Thinking Fast and Slow,* 7.

situations where deductive methods are epistemically superior to inductive methods in supporting mathematics claims. Individuals who have received serious mathematical training are therefore likely to have a preference for deductive methods over inductive methods, feeling that nothing short of proof is acceptable. Thus making use of Algorithm 3.12 is felt to be a better method than employing Algorithm 3.13, which is treated with suspicion – even though in this particular case the chances of an error arising due to the randomised nature of the algorithm was utterly negligible compared to errors caused by other factors, and using this as the sole basis for judgement is therefore irrational.

3.vii. Conclusion

Some Monte Carlo algorithms are such that errors arising due to their inherently randomised nature seem negligible when deciding whether to believe the results of the algorithm, because the probability of such errors are manifestly tiny in comparison to those of other sources, such as human or hardware failure. In the next chapter I shall carefully argue that in some cases the overall probability of error – when these other, external sources of error are taken into account – can be smaller for such non-deductive algorithms than for their deductive equivalents. In particular, we will show that this was true for the experiment conducted in the previous section, where the integers 48,140,819 and 66,998,713 were produced by deductive and Monte Carlo techniques respectively.

4. Two Kinds of Error

In this chapter, we will see that for the experiment performed in the last chapter the Monte Carlo algorithm provided a more reliable means of arriving at results than the Trial Division Algorithm – a deductive rival that which mathematicians would currently regard as acceptable. I argue that mathematicians are under such circumstances rationally obliged to yield Private Acceptance to the results of these algorithms. I then consider whether Monte Carlo algorithms are suitable for use in the context of Public Acceptance. I further argue that the four Practical Virtues of mathematical practice – Permanence, Reliability, Consensus, and Autonomy – would not be affected by this revision to established practice.

4.i. Computer Errors

In this section, we will review the experiment done in the previous chapter, whereby we gave two methods for finding a prime number. These were Algorithm 3.12, the Trial Division Algorithm, and Algorithm 3.13, which used the Rabin-Miller test. We consider various sources of error that could have arisen on our route to arriving at the two purportedly prime numbers $48{,}140{,}819$ and $66{,}998{,}713$, concluding that the latter method was more reliable.

Firstly, I could have made a mistake in programming the algorithms: that is, the programs I have actually produced may not faithfully embody the correct set of instructions to be carried out in either case. In this instance, my implementation – which may be far from the most efficient – of the Rabin-Miller Algorithm used some 128 lines of code, none of which was particularly complicated. The Trial Division Algorithm used somewhat less code; around 100 lines. Hence, we may take it that the chances of an error arising here are about the same. This case is somewhat atypical, because of the unusual simplicity of the Trial Division Algorithm: most modern algorithms for finding prime numbers are very complex indeed in comparison to the Rabin-Miller Algorithm, and hence programming errors are in general far more likely in the deductive case.

There are also other possibilities for human error in using the algorithm. I had to read the answer off the compiler output and transcribe it correctly into this manuscript. When I reminded the reader of our two candidates in the opening paragraph of this section, I also wrote the numbers down from memory. We might hope these kinds of errors are negligible, but if we could put a precise number on the probability of their occurrence it would surely dwarf the number $\varepsilon = 10^{-10,000}$, our upper bound for the error introduced by the Monte Carlo nature of Algorithm 3.13.

Because the chances of programming and other human errors occurring will be broadly equivalent in both cases, we disregard these and focus on error introduced when the computer runs the algorithm, assuming that it has been programmed correctly. Firstly, some useful terminology.

Internal Errors

These are errors introduced by the inherently probabilistic nature of Monte Carlo algorithms, for which it is possible to receive an incorrect answer even if the algorithm is put into practice perfectly.

Implementation Errors

This will be used as a general term for the various kinds of software and hardware failures that arise when an algorithm is run on a computer.

So far, we know that Algorithm 3.12 is free from Internal Errors and further that the probability of an Internal Error when running Algorithm 3.13 is less than $\varepsilon = 10^{-10,000}$. We now enquire into the relative probabilities of Implementation Errors in both cases and compare them to this probability of Internal Error.

Modern computers are impressive pieces of equipment, but they have not yet attained the reliability of HAL9000, the mischievous computer from Stanley Kubrick's film *2001: A Space Odyssey*, who boasts that 'The 9000 series is the most reliable computer ever made. No 9000 computer has ever made a mistake or distorted information. We are all, by any practical definition of the words, foolproof and incapable of error.' Just like humans, computers can only store and perform operations on data with less than perfect fidelity: 'The computer may slip a pulse, its voltages may drop, it may be communicated with over a noisy channel.'[160]

In a paper called 'Computer Programming for Accuracy', Mike Yohe lists 38 types of errors that may occur in running a computer program, including so-called 'soft errors'[161] that are not the fault of the program itself. These types of error are grouped under the seven categories of 'errors due to hardware limitations, errors due to software limitations, error due to hardware failure, errors due to software failure, errors due to program failure, errors due to faulty operation, errors due to inadequate planning'.

Yohe introduces quantitative methods for giving upper bounds for the effects of these types of error, and practical suggestions for avoiding some of them. The robustness of systems can also be increased with error-correction software, but this has limitations and is itself vulnerable to error. Computer hardware is susceptible to numerous unpredictable, often unpreventable or undetectable environmental conditions, such as the influence of nearby terminals or even cosmic rays.[162] And the random decay of chip material releases alpha particles that can change the state

[160] Philip Davis, "Fidelity in Mathematical Discourse: Is One and One Really Two?", *The American Mathematical Monthly*, 79 (1972): 256.
[161] J. M. 'Mike' Yohe, "Computer Programming for Accuracy", *Proceedings of the 1968 Army Numerical Analysis Conference, U. S. Army Research Office, Durham, North Carolina* (1968). Quoted in Davis, "Fidelity in Mathematical Discourse", 257.
[162] James Ziegler and William Langford, "Effect of Cosmic Rays on Computer Memories", *Science* 16 (1979), 776-788.

of a memory cell to a different value. So the possibility of errors arising is never eliminated entirely.

We divide the Implementation Errors that may have occurred when running either algorithm into two types: errors caused by the computer's memory undergoing distortion over time; and errors made when the computer performs specific operations, such as carrying out a calculation, or retrieving, storing or modifying the value of a variable.

The first kind of error can be expected to occur randomly and at a uniform rate, so we may model it using a Poisson distribution. The latter kind has a probability given by the Binomial distribution. This is known to be approximated by a Poisson distribution when the number of operations is sufficiently large, which is certainly the case here. So overall both kinds of error can reasonably be modeled with a Poisson distribution. We can therefore expect the number of errors of each kind to be proportional to the running time of the algorithms, which for simplicity we will regard as proportional to its computational complexity.

To show that p is prime using the Trial Division Algorithm, we need to check through $\lfloor\sqrt{\mathrm{p}}\rfloor - 1$ potential divisors d. For each of these, we need to compute $p \bmod d$, which was achieved by subtracting d from p until the residue was found, requiring around p/d operations.[163] Hence, the total number of operations is approximately $\frac{p}{2} + \frac{p}{3} + \cdots + \frac{p}{\lfloor\sqrt{p}\rfloor} \cong p\ln\sqrt{p}$. We therefore take the total number of operations to evaluate a candidate to be $\mathrm{O}(2^n \ln(2^{n/2}))$, where n is the number of binary digits of the required output.

The Rabin-Miller Algorithm is much faster, however. Using the repeated squaring method of exponentiation, the complexity of evaluating a single candidate is here $\mathrm{O}(k\ln^3(2^n))$, where k is the maximum number of values of a we test as witnesses, and n is the number of binary digits of a candidate.[164]

As around $1/\ln(2^n)$ of numbers in the set from which candidates are selected are prime, we will need on average to try approximately $\ln(2^n)$ candidates before finding a prime. Disregarding the impact of composite numbers passing all the tests in the Rabin-Miller case, this gives us estimates for the overall complexities of each algorithm as $\mathrm{O}(2^n \ln(2^{n/2})\ln(2^n))$ and $\mathrm{O}(k(\ln^4 2^n))$ for the Trial Division and Rabin-Miller algorithms respectively.

Substituting in the values $k = 16{,}613$ and $n = 26$, we see that the approximate total numbers of operations are of order 10^{10} and 10^9 respectively. Hence, more operations will be needed for the Trial Division Algorithm, even though in hindsight it must be said that we were rather extravagant with the number of Rabin-Miller tests run.

---

[163] There are much better ways: see for example Crandall and Pomerance, *Prime Numbers*, 414, algorithm 9.2.10, who write 'The complexity is the same as a size-*N* multiply', where *N* is the base. However, this method is what I used for the experiment.

[164] Gary L. Mullen and Daniel Panario, *Handbook of Finite Fields,* (Florida: CRC Press, 2003), 347. We write $f(x) = \mathrm{O}(g(x))$ iff $\exists\, M, x_0 \in \mathbb{R}^+$ with $|f(x)| \leq M|g(x)|$ for all $x \geq x_0$.

These calculations show that we should expect more of both memory-distortion and operational errors when running Algorithm 3.12 than Algorithm 3.13. This does not resolve the issue of reliability, however. The occurrence of either kind of error is never a guarantee that the algorithm will give us a composite rather than a prime number, so for evaluating the chances of this happening we must also see how these errors will actually affect the overall output of the algorithm. This will require more detailed consideration of how each algorithm works.

We first consider the impact of errors that might have been induced when the computer performed specific operations, beginning with the Trial Division Algorithm. Once a candidate is selected, it is evaluated as either prime or composite. Two kinds of error are possible here. Either the program could falsely assert a composite number to be prime, or it could falsely assert a prime number to be composite. The latter kind of case is of no concern to us here: in this instance, the compiler would simply discard the number and look for another candidate. The only circumstance under the computer gives an erroneous output is if a composite candidate is falsely deemed to be prime.

Let us consider how this might occur by using the composite number $N = 799 = 17\times47$ as an example. The compiler will attempt to divide N by the numbers $2, 3, 4, \dots, 27, 28$, as $\sqrt{799} \cong 28.267$. Suppose further that the algorithm makes a mistake when computing, for example, $799 \bmod 13$. If it incorrectly computes this remainder as 0 instead of 6 then it will correctly identify 799 as composite (although for the wrong reasons). It will then discard it and look for a new candidate. If it computes this remainder as a different non-zero number, other than the correct value 6, then the algorithm will still discover that 799 is composite when it later successfully divides it by 17, and the error will again have been of no consequence. In fact, the only circumstances under which this type of error will lead to the compiler falsely asserting that 799 is prime is if it incorrectly calculates the value of $799 \bmod 17$.

More generally, the Trial Division Algorithm will only incorrectly assert a composite number N to be prime because of a mistake in calculating the remainders as it computes $n \bmod d$ incorrectly for every number d such that $d|N$ and $d \leq \sqrt{N}$. This probability depends on the specific factorisation of N. As N can be expected to have around $ln(lnN)$ factors, then we may expect the probability of this occurring to be ${\varepsilon_1}^{\ln(\ln(N)/2}$, where ε_1 is the probability of a particular computation of a residue going wrong.[165] For our experiment, this value was around ${\varepsilon_1}^{1.45}$.

We now consider parallel kinds of Implementation Error when running the Rabin-Miller Algorithm. Let ε_2 be the probability that the compiler makes a mistake by failing to notice that a given number is a witness to the compositeness of n. In general we know that there are at least $3(N-1)/4$ witnesses for composite N, so we should expect to pick around $3k/4$ witnesses in total when running our k tests. But this time, we need the compiler to incorrectly fail to identify every single one of these as witnesses. We may therefore estimate that the chances of this having

[165] In deriving this result, I made use of the Erdős–Kac_theorem.

happened in our experiment were less than $\varepsilon_2{}^{12459}$. Whatever the correct empirical values of ε_1 and ε_2 turn out to be, it is clear that the overall chances of an operational error causing the Rabin-Miller Algorithm to output a composite number are far less than in the deductive case.

As became apparent in Chapter 1, the very strength of proof within the classical tradition – its cumulative, deductive structure – becomes a kind of weakness when informed by the new computer-based perspective where proofs are much longer and the laws of information theory come into effect, because any gap in an argument can render a proof invalid. In contrast, we can now see that because IMC algorithms work by completing a large number of independent tests, if there is an error or glitch that is localised to the performance of just one of these tests then the this actually matters very little. As with Internal Errors, the chances of such Implementation Errors actually causing an incorrect output decreases exponentially with the number of tests.

Data corruptions due to environmental stimuli such as a cosmic ray changing the state of a memory cell may affect the running of either algorithm in a variety of unpredictable ways. Some might cause an effect of a global nature, whereas others might be less significant. For instance, the Trial Division Algorithm requires a variable that enables the compiler to keep its place in the sequence of divisors; this will start from 2 and increase up to 28 in the case just illustrated. However, supposing after the third iteration a cosmic ray causes its value to change from 4 to 36 unexpectedly; then the algorithm will terminate here and not go on to discover that $17|799$. It will then falsely conclude that 799 is prime, having not found a factor by the end of the process.

The Rabin-Miller Algorithm also needs a variable that counts the number of tests that have been run, and so interference with the computer's memory could alter its value and cause the algorithm to terminate prematurely. However, most of the time when this occurs, a large number of independent tests would have already been run by this point. Hence the threat of such errors – in themselves likely to be more numerous, due to the longer running time – is greater for Algorithm 3.12. In conclusion, because Algorithm 3.13 takes much less time to run and tends to deal with errors in a more robust way, we may take it that the probability of an Implementation Error is less than for Algorithm 3.12.

Now, let us compare these Implementation Errors to the probability of an Internal Error in Algorithm 3.13. The age of the universe is currently estimated to be around 4×10^{17} seconds. Suppose we proceed to run our Monte Carlo algorithm 250 times per second on a supercomputer. We would on average have to continue doing this for a hundred million billion billion billion … billion (with 'billion' written 1108 times) times the entire age of the universe before we encountered a single Internal Error. It is clear that Implementation Errors must occur in both algorithms at a much greater frequency than this. But because we can reasonably approximate the total probability of error as the sum of the probability of Implementation and Internal Errors, we see that the overall probability of error associated with Algorithm 3.13 is in fact less than that for Algorithm 3.12.

4.ii. Knowledge and Epistemic Externalism

The conclusions of the previous section suggest that contrary to my initial intuitive reaction we would have been wiser to give the daemon the number 66,998,713 in the context of the thought experiment described in Section 3.vi. In the next two sections, we consider a philosophical objection to forming beliefs on the basis of the output of a Monte Carlo algorithm. The objection is that use of such algorithms can never lead to knowledge of the corresponding conclusions, where 'knowledge' is understood in an infallibilist sense. As knowledge is taken to be of central epistemic importance, this may provide mathematicians with sufficient reasons to reject Monte Carlo algorithms in the context of Private Acceptance.

Consider the kind of intuition appealed to in lottery cases, as discussed by Hawthorne and others.[166] An individual buys a ticket for the UK national lottery, giving him a 1 in ${}^{49}C_6 = 13{,}983{,}816$ chance of winning the jackpot by correctly selecting the 6 winning numbers from a possible choice of 49. It is thus very unlikely he will win. However, a logical problem seems to arise if we assert that he *knows* he will not win. For it is plausible to suggest that however we are to understand propositional knowledge, it should be closed under conjunction, in the sense that if a subject knows propositions A and B and is familiar with the rules of logic and validly infers $(A \wedge B)$ then he or she knows this proposition too.

Suppose then that our first individual knows his ticket will not win. In this case, a second, very wealthy individual could buy all 13,983,816 tickets for the lottery and choose every possible combination of numbers. By symmetry with our first speculator, he knows for each ticket that that ticket will not win. He therefore knows that none of them will win, as knowledge is closed under conjunction. But this is preposterous, as in fact he knows that one of them must win.

Based on this kind of example one might conclude that knowledge should be understood in an infallibilist sense, at least in certain kinds of cases where this line of reasoning applies. There are of course various cogent objections to this move in the epistemological literature, but let us concede this point to the infallibilist for the sake of mounting the challenge against Monte Carlo algorithms. If infallibilism turns out to be untenable, then as we shall see the argument cannot get going.

In cases similar to the lottery example, then, the grounds that a subject possesses as the basis for his or her belief do not guarantee that the belief in question is true – in this case, the belief we will not win – and as a result we say that the belief is not knowledge, even though it is very likely to be true. But a subject who forms beliefs on the basis of a Monte Carlo algorithm such as Algorithm 3.13 is clearly in a situation that is similar in the relevant respects. So if the infallibilist conception of knowledge does applies here too, then such a belief – for example, the belief that 66,998,713 was prime that I formed in the last chapter – can therefore never qualify as knowledge on the basis of these algorithms alone.

We might initially think that an individual who gained a belief from a deterministic program – and again, we may take as an example our belief that

[166] John Hawthorne, *Knowledge and Lotteries* (New York: Oxford University Press, 2004).

48,140,819 is prime – would likewise also not be in possession of knowledge. For as we saw in the previous section, such programs are subject to a variety of hardware, software and human errors. So the passage to belief can also be associated with some positive probability of failure. However, I shall now present a line of thought that argues this is not so (continuing to assume the infallibilist conception of knowledge throughout, again for the sake of argument).

Firstly, we make a distinction between the grounds that a belief is based on, and the access we have to those grounds. Consequently, when we believe a proposition on the basis of some such grounds, there are two possible sources of error. Either the grounds we have for believing the proposition do not guarantee its truth, and this is one of the cases where the grounds obtain but the conclusion does not; or, whilst the grounds we think we have do indeed guarantee the truth of the proposition, we are in fact mistaken about having the grounds we think we have.

Now, consider an analogy with visual perception. Sarah, who is at a wildlife enclosure, looks out across a field and sees what appears to her to be, and in fact is, a zebra. She thus forms a true belief that there is a zebra in the field. Meanwhile, Jeff – who is at an entirely different wildlife enclosure in a different part of the country – looks out across a field and also sees what again appears to him to be a zebra. However, this time it is in fact not a zebra, but rather a cleverly painted mule.[167]

There is of course a well-known skeptical argument for the conclusion that Sarah's true belief that there is a zebra in the field does not constitute knowledge, because the case of Jeff clearly shows that having the experience of seeing what appears to be a zebra is not a logical guarantee of the zebra's actually being there. The case is thus construed as similar in the relevant respects to buying a lottery ticket, even though this time we are unable to attach so precise a probability to the evidence misleading us. Yet this kind of skeptical argument is now found to be unconvincing by many philosophers, because we can characterise Sarah's grounds for belief in a different way.

Consider first the logic of how truth conditions attach to verbs relating to perception, such as 'see'. There are two kinds of accounts. According to one sort of view, 'seeing a zebra' involves only being aware of certain mental entities that give us the subjective impression of seeing a zebra. On this account, both Jeff and Sarah are seeing a zebra. A second type of account understands seeing as *successive*: that is, that seeing constitutively involving us standing in a certain causal relation to the seen object. For it to count as a case of seeing in particular, the seen object must therefore actually exist. For theories of seeing of this kind there is thus an important logical difference between the two cases: despite the subjective indistinguishability of Sarah and Jeff's experiences, only Sarah but not Jeff really sees a zebra.

We need not discuss which of these two types of theories is correct – we simply assume the latter kind of account is true for the purposes of the argument (it is far

[167] The example is of course borrowed from Dretske, though he puts it to quite different use. Fred Dretske, *Knowledge and the Flow of Information* (Cambridge: The MIT Press, 1981).

more popular in any case). Now, if we also assume that Sarah's grounds for believing there is a zebra are understood to be that she specifically *sees* a zebra and not merely that she *appears to see* a zebra, then she does have grounds which logically entail the truth of her belief. For as we have already said, if it is to count as a case of her seeing a zebra (in the successive sense) then it is necessary for the zebra to actually exist.

The focus of the discussion now shifts to the second of the two ways in which an error can arise: whether Sarah really does have the grounds she thinks she does. We can agree that from her perspective it is *possible* that she sees a cleverly painted mule rather than a zebra, in the sense that this would have been subjectively indistinguishable from what actually occurred. However, this would require her world to be very different in the relevant respects (assuming here that there are no cleverly painted mules in the surrounding area). Hence we may take it that this possibility does not undermine her claim to knowledge. After all, her mode of access to her evidence is reliable – we may suppose she enjoys good eyesight and is here subject to normal viewing conditions – and the acquisition of her belief through perception no way involves luck or a violation of epistemic rationality (again, skeptical readers are invited to go along with this thought merely for the purposes of the argument).

Now, suppose James – a mathematician – discovers a new result in geometry by spotting an invariant using *Apollonius* and is then able to produce a proof for his claim. When asked why he believes his theorem to be true, he responds that he knows it is true because he has proved it. As we have already noted, the term 'proof' denotes a successive concept. It is therefore natural to understand the verb 'prove' in a successive sense, analogous to the second type of account of seeing. The argument he has discovered is in fact sound, he is in general fairly reliable – though perhaps not infallible – at telling when this is the case, and he was in no way lucky in constructing his argument. Hence, we may say that James *knows* the content of the theorem, even if he had upon occasion been unable to tell that a similar-looking mathematical argument was in fact flawed in past instances.

Lastly, we transfer these ideas to the case at hand. Suppose Susan correctly programs Algorithm 3.12 onto a computer and then runs the algorithm with no Implementation Errors occurring, thus coming to the true belief that a certain number is prime. By analogy with the previous two cases, we may characterise the grounds of her belief as that a series of computer operations have been performed which can easily be converted into the presentation of a proof that the outputted number is prime. In the skillful use of the computer to determine this she has made use of a kind of faculty that is analogous to the case of perception here, though less direct, and we may again understand the operation of this faculty in a successive sense. Assuming that she is at least a fairly reliable programmer, and the machine the program is implemented on is tolerably robust, our subject has gained access to conclusive grounds for her belief through a reliable method that involves no violation of epistemic rationality. The belief thus constitutes knowledge.

In contrast, an individual running only Algorithm 3.13 can never have knowledge that the outputted number is prime, even if the algorithm is programmed correctly and implemented without incident, because his own characterisation of his

grounds could be no stronger than the statement of the inconclusive fact that a Rabin-Miller program outputted that number. There are no conclusive grounds available that he is in a position to gain access to using only this program (it is immaterial whether the number is actually prime or not). So only a deterministic algorithm and not a Monte Carlo algorithm can yield knowledge. Though we may perhaps now know that 48,140,819 is prime, the same cannot be said for 66,998,713. And clearly, the same argument applies in similar cases.

4.iii. A Pragmatic Approach to Epistemic Concepts

The conclusions of the above argument – that only the less reliable source of belief can enable us to have knowledge – gives us excellent grounds to think that the infallibilist conception of knowledge we have partially articulated must be in some respects flawed. But let us take it that the argument given above is correct on its own terms: that knowledge is to be understood in an infallibilist sense in these kinds of cases, that a successive account of the relevant evidential relations holds, and hence only Algorithm 3.12 but not Algorithm 3.13 can furnish us with knowledge that a particular number is prime. How does this conclusion affect our feelings about the results of the experiment performed in the previous chapter? The reader is again invited to imagine that he or she has an urgent need to supply a prime number in this range for some particularly important application, and to revise their earlier choice of either 48,140,819 or 66,998,713 if necessary.

If we take the application of the concept of knowledge – as understood in the infallibilist sense – as primary here, the conclusions of the above argument provide us with a principled means of choosing. However, we in fact have a variety of analytic tools at our disposal: there are other evaluative epistemic concepts such as justification, grounds, evidence, and – importantly – the concept of reliability itself. We could even simply decide to think in terms of a knowledge-like concept that operates under a logic that is fallibilist. As rational agents, we are able to reflect on the effectiveness of using a particular framework of concepts when formulating and pursuing our goals.

If our primary concern is with finding the daemon a prime number, then, we should not prejudice our account by making an *a priori* assumption that one of these concepts in particular, such as knowledge as understood by the infallibilist, will give us the most illuminating terms in which to understand our situation or best help us assess the possible courses of action open to us. We are not aiming to manoeuvre ourselves into the extension of any fixed epistemic concept as such, but rather are primarily concerned with a definite practical goal. For any given concept we possess, we can always enquire as to its importance and to question continuing to yield it the role within our practical deliberations that it may have enjoyed hitherto. So we would be prudent to adopt a pragmatic attitude to epistemic concepts here, as opposed to insisting on going in 'knowledge first' or some other approach.

Let us think more carefully about the form that the two options are presented to us in. Regardless of whether either or both of the programs did indeed run correctly, the choice as it now appears to us, from the first person perspective, is whether to

endorse the outcome of Algorithm 3.12 and give 48,140,819 as our answer, or to endorse Algorithm 3.13 and give 66,998,713. Notwithstanding how the logical analysis of our grounds would appear if the case were described in the abstract, if all the facts were known, this is how the options are presented to us. We have no direct access to the actual sequence of calculations that have taken place; the detailed reasoning process has become externalised into the computer. And the epistemic concepts we make use of in making our decision must also be such that we can successfully deploy them here and now, within our particular situated position, to help us decide.

Now, clearly the concept of knowledge enjoys a central role in both our practical and theoretical lives. Yet when we try to employ it in making a decision here, a problem arises. Let us deliberate: we may take it that the method by which we found 48,140,819 would be such as to yield knowledge, assuming the algorithm was programmed correctly and ran with no Implementation Errors. But *did* this happen, and *has* it yielded us knowledge in this case? *Do* we now know that 48,140,819 is prime? We have the conditional 'if the algorithm ran correctly, we know; if not, not', but no direct insight into whether its antecedent is true. Compare: the wife of an analytic philosopher asks him whether she should take an umbrella to work that day. He pauses, then carefully answers: 'If it will rain today, you should; if not, not.' This is true enough, but it does not help her decide what to do. An enquiry into the logic of which concepts apply or do not apply in each possible situation is simply insufficient if we are not sure which of these situations actually pertains.

Fortunately, we have another analytic tool that we can deploy operationally here: the concept of reliability. This time we are able to apply the concept successfully: indeed, we have done this already in Section 4.i. And knowing the distribution of error seems to give us all the information we need to make a decision. We suggest the following as a general principle of epistemic rationality:

Reliability Principle

When seeking an answer to a fixed question, always give preference to a source of beliefs known to be more reliable over a less reliable source, assuming both are freely available.

Applying this to the case at hand, there were two sources for gaining a new belief: running Algorithm 3.12 on my laptop, and running Algorithm 3.13 on the same laptop. Interpreting reliability here in the natural way, as the probability that either program will give a correct output, we have the following table:

	Ideal World	**Real World**
Algorithm 3.12	1	$1 - p_{imp}$
Algorithm 3.13	$1 - q_{int}$	$1 - q_{int} - q_{imp}$

Table 4.1, showing reliabilities. p_{imp} and q_{imp} are the Implementation Errors, q_{int} is the Internal Error of Algorithm 3.13, and we have argued that $q_{int} + q_{imp} < p_{imp}$.

From the perspective of current mathematical practice, the crucial difference is in the left-hand column: Algorithm 3.12 works better in principle, even if this is not true in practice. But from a normative epistemic perspective, this is puzzling. We *know* that we live in the real world, where our physical systems yield only fallible means of processing information: therefore the right-hand column is the one to which we should attend. An important part of the argument of the previous section was that Susan had no special reason to doubt that her method was implemented in a reliable way. But as we are supposing that we care deeply about the outcome, there is no reason for us to now restrict ourselves to considering only those possibilities that we are *obliged* to take into account in order for us to satisfy the conditions governing the ascription of propositional knowledge. Rather, we can better ask: which possibilities are we now *able* to take into account that will lead us to making the most informed choice, and hence to success in our practical goal of supplying a prime number?

It is clear then that if we can factor Implementation Errors into our analysis then surely it would be wise to do so, all other things considered. We should therefore give the answer $66{,}998{,}713$ to the daemon, because this number was derived through a process that was more reliable (the reader is again asked to reassess at this stage which number they would intuitively prefer to give to the daemon, assuming that very high stakes are in play).

We now generalize our conclusions from the particular case of our experiment to the rationality of Private Acceptance as such in this kind of situation. We saw that the first-person perspective arose quite naturally when we put the algorithms into practice to attain definite results, and this will also be the case for any individual mathematician in similar circumstances. Hence instead of looking only at the abstract, logical relations between our received epistemic concepts, we may follow Descartes and aim to model quite generally the position of first-person enquiry. Therefore, wherever mathematicians make use of the results of deductive algorithms in the context of Private Acceptance, these same mathematicians would be irrational not to believe the results of equivalent Monte Carlo randomised algorithms that are known to be more reliable overall.

Lastly, we make some more general observations about the relative reliability of deductive and Monte Carlo algorithms in the context of finding prime numbers. Admittedly, the Trial Division Algorithm could have been implemented far more efficiently for Algorithm 3.12: the number of operations would be dramatically reduced not only by using a faster method for computing residues but also by considering only the prime numbers less than $\sqrt{n}$ as possible divisors. Further reductions to the number of divisors are also possible: for example, by employing the concept of a 'wheel'.[168] These improvements may increase the impact of some kinds of Implementation Errors, however: whenever we increase efficiency by eliminating redundancy we tend to make our systems less robust.

Moreover, even the best deductive algorithms are much slower than the Rabin-Miller Algorithm, which can itself be improved to $\widetilde{\mathrm{O}}(k \ln^2 n)$ using Fast Fourier

[168] Crandall and Pomerance, *Prime Numbers*, 111.

Transform-based (FFT) multiplication.[169]Consider, for example, the Agrawal-Kayal-Saxena (AKS) primality test, one of the most efficient modern algorithms for determining whether a number is prime.[170] The complexity of the original algorithm was $\widetilde{O}(\ln^{12} n)$, although Carl Pomerance and H. W. Lenstra[171] later came up with a version whose complexity is only $\widetilde{O}(\ln^{6} n)$. It is clear when looking at larger numbers the Rabin-Miller Algorithm is far superior in this respect: for instance, in verifying a candidate with 200 binary digits by running $20{,}000$ tests, say, the modified Rabin-Miller Algorithm will need to perform some small multiple of 3×10^{8} operations, whereas the order for the modified AKS test will be roughly of magnitude 7×10^{12}. This is around twenty thousand times as many, though we have made an even more extravagant choice for the number of tests.

In general, then, we may expect more Implementation Errors to occur when running deductive programs than with their Monte Carlo rivals. Moreover, these errors are sufficiently likely to render the possibility of an Internal Error negligible in comparison. So in using Monte Carlo algorithms we are likely to make fewer errors overall, compared with relying on the deductive algorithms mathematicians have traditionally preferred.

4.iv. Human Errors

> 'In all demonstrative sciences the rules are certain and infallible; but when we apply them, our fallible and uncertain faculties are very apt to depart from them, and fall into error… Our reason must be consider'd as a kind of cause, of which truth is the natural effect; but such-a-one as by the irruption of other causes, and by the inconstancy of our mental powers, may frequently be prevented.' David Hume[172]

In this section, we take another brief detour by continuing the discussion of Section 1.v about the Reliability of mathematical discourse, here insofar as it is produced by human mathematicians. Although impressive, we shall soon see that the abilities of mathematicians in this respect are less than perfect. The relevance of this observation – soon to be supported by philosophical and historical argumentation – is to block another argument against Monte Carlo algorithms. A critic might recognise their superiority to deductive algorithms, but simply maintain that deductive algorithms should also be abandoned too. In this section we will see that we should not worry too much about the fallibility of computers. The production of new mathematics has always been attended with some chance of error, and human error is by far the more likely cause of a mistake.

We have said that 'proof' is to be used only in a successive sense: that is, a sound deductive argument for a particular claim. Hence by definition proofs establish

---

[169] The 'soft-O' notation $f(x) = \widetilde{O}\big(g(x)\big)$ means that $\exists\, k \in \mathbb{R}$ with $f(x) = O(g(x)\log^{k}(g(x)))$.

[170] Agrawal Manindra, Neeraj Kayal and Nitin Saxena, "PRIMES is in P", *Annals of Mathematics* 160 (2004): 781-793.

[171] "Primality Testing with Gaussian Periods", accessed 13th August 2015, https://math.dartmouth.edu/~carlp/aks041411.pdf

[172] David Hume, *A Treatise of Human Nature* (Oxford: Oxford University Press, 2000), 121.

their conclusions with no possibility of error, and it makes no sense to say 'she found an error in her proof', but we should rather say something like 'what she initially took to be a proof, later turned out not to be'. However, as also made clear above, even if mathematicians unanimously adopt the *policy* of aiming to present proofs for all of their published writings, mistakes can and do creep into the literature. This is because mathematicians fall short of perfection in their ability to construct discourses that actually do give expression to proofs, as opposed to merely giving the appearance of doing so, and to check whether this is the case for a discourse written by another mathematician.

Teaching experience discloses that mistakes occur frequently in the written work of most mathematics students, even by the time they are taking undergraduate degrees. Numbers and symbols are garbled during algebraic manipulations, expressions are copied incorrectly, theorems are applied when the conditions of application are not all met, elements of sets are counted multiple times, exceptional cases to which the main argument does not apply are ignored. A large number of errors occur from failure to treat minus signs with sufficient care: when multiplying out brackets, for example. Several years ago I invigilated an hour-long logic exam sat by eighty students, where the papers were marked anonymously and then correlated using seven-character 'candidate numbers', which the students were asked to copy onto the front of their exam papers. Three students, who all went on to score fairly highly in the logic exam, copied their candidate numbers incorrectly – even though the task was simple, there was ample time to check, and their passing the module was potentially at stake. And it is clear that the probability of there being at least one transcription error would increase with the number of students considered.

There are steps that can be taken to avoid such errors. With numerical questions it is good practice to construct a Fermi-style approximation for how big the answer is likely to be, either mentally or on paper, to serve as a kind of check. Many errors are caught after examining an answer more carefully because it deviates from what is expected. However, it is surprising how little some students flinch when an obviously incorrect answer is found. Some years ago a young student of mine was calculating the original salary earned by someone who now received £33,000 a year after a 10% increase; after dividing by 110 to reach 1% of the initial salary they forgot to multiply by 100, and confidently wrote down a figure of just £300 a year! The frequency of error also varies with our level of concentration, which one can take steps to enhance, such as working in a quiet environment, and drinking plenty of water. But even with these aids it seems impossible to eliminate the possibility of error entirely.

Moving from teaching to research contexts, the frequency of these mistakes does tend to diminish with years of mathematical training and experience. There may be many reasons for this. Perhaps it is partly due to an increase in cognitive power that enables professional mathematicians to hold larger chunks of the argument in their heads, and thus rely less on visual recaps. Perhaps it is due to acquiring habits that tend to reduce errors, or learning where certain natural pitfalls are in a variety of kinds of argument and how to avoid them. A greater depth of understanding also helps: we are far less likely to misremember the value of $\sin(\pi/3)$ as ½ instead of $\sqrt{3}/2$ if we visualise it geometrically, as the y-coordinate of a point

with polar co-ordinates $(1, \pi/3)$. Lastly, there may be a number of overlapping checks in place that must all fail for error to occur unnoticed. A number theorist adding 125314 together with 141351 would immediately know they had made a mistake if the answer was an even number because of their intuitive grasp of parity, and likewise if the product of these two numbers turned out to be odd.

Impressive as the abilities of a developed mathematician may be, however, they are nevertheless fallible in performing calculations. As Philip Davis points out in an entertaining and provocative paper entitled 'Fidelity in Mathematical Discourse: Is One and One Really Two?', all mathematics is expressed and communicated through the use of discrete symbols, which we interact with through a physical trace, such as a blob of ink, or a vibration in the air. These cannot be created, recognised, reproduced and concatenated with perfect fidelity, because human mathematicians – like computers – are fallible when considered as physical symbol-processing systems.

This fact may have been easier to ignore where proofs were generally fairly short and information-theoretic considerations did not arise, so that our ability to manipulate symbolic expressions could be considered absolute without obvious contradiction. But as we saw in Chapter 1, some proofs are so long that this is now no longer the case. Describing day-to-day mathematical practice, Davis and Hersh write that making errors 'happens to the best of us every day of the week. When the error is pointed out, one recognizes it as an error and acknowledges it. This kind of situation is dealt with routinely.'[173] Likewise, philosopher Jody Azzouni writes: 'It's a robust part of mathematical practice that mistakes are found and corrected.'[174]

These observations add weight to De Milo, Perlis and Lipton's claim in an influential paper that the Reliability of the mathematical literature is secured not by individual mathematicians but because proofs are always positioned within a network of discoverers, checkers, reviewers, editors, users, communicators, simplifiers and generalisers.[175] Once a proof is discovered it is generally articulated to colleagues, who then break it down and internalise it, producing their own versions and making simplifications. If it passes these initial informal tests, then the referees take a more careful look. If it is a significant result then once published it will be connected to a wide variety of other work, and then subjected to scrutiny for many years: every time the proof is articulated to new students, for example. The Reliability of the most central parts of mathematics is thus is achieved only via this social and institutional process, as was also well understood by Hume:

> 'There is no Algebraist nor Mathematician so expert in his science, as to place entire confidence in any truth immediately upon his discovery of it, or regard it as any thing, but a mere probability. Every time he runs over his proofs, his confidence encreases; but still more by the approbation of his

[173] Davis, Hersh and Marchisotto, *The Mathematical Experience*, 61.
[174] Jodi Azzouni, "How and Why Mathematics is Unique as a Social Practice", 207.
[175] Richard De Millo, Richard Lipton and Alan Perlis, "Social Processes and Proofs of Theorems and Programs".

> friends; and is rais'd to its utmost perfection by the universal assent and applauses of the learned world.'[176]

So far we have discussed common but somewhat superficial algebraic or technical errors, most of which can easily be patched up, and merely typographical mistakes that constitute only the isolated and relatively unproblematic misprinting of a single symbol. Yet enquiry into the history of mathematics shows us that mathematicians also commit more serious conceptual mistakes that can threaten the validity of an entire proof strategy. These include some of the highest profile conjectures of modern mathematics: we have already noted serious mistakes in the initial versions of the Enormous Theorem papers and Kempe's flawed argument for the Four Colour Theorem.

Davis estimates that an error of 'international significance' – that is, the 'conjunction of a mathematician of great reputation and a problem of great notoriety' – occurs every 20 years or so.[177] Moreover, for these deeper conceptual errors the social, probabilistic nature aspect of the checking network can be less effective and the mistake can stay undiscovered for a long time: Kempe's argument stood for 11 years before Heawood discovered his error. Once Heawood pointed it out then it was easy enough to see, but spotting it required the use of imagination where it was not obviously required. We close this section with more concrete examples of mistakes made by mathematicians.

- Euler believed he had proved that for any function f of two real variables the order of taking partial derivatives did not matter; that is that the equation $\frac{\partial^2 f(x,y)}{\partial x \partial y} = \frac{\partial^2 f(x,y)}{\partial y \partial x}$ always holds. His argument was flawed, and a counterexample was given by Schwarz in 1873.[178]

- Given a right-angled triangle, the Malfatti circles are the unique trio of circles that are each tangent to the other two and to a pair of sides. In 1803, Gian Francesco Malfatti claimed to have proved that of all the different ways of inscribing three non-overlapping circles into a right-angled triangle, the Malfatti triangles maximize the total area enclosed by the circles.[179] Lob and Richmond disproved the conjecture in 1930,[180] and in 1967 Goldberg used their constructive procedure to show that Malfatti's solution is *never* optimal.[181]

---

[176] Hume, *Treatise*, 121.
[177] Davis, "Fidelity in Mathematical discourse", 262
[178] Steven Engleman, *Families of Curves and the Origins of Partial Differentiation* (Amsterdam: Elsevier, 2000), 9-11. The content of Euler's conjecture is not entirely clear because the function concept had yet to receive a clear formulation: see Section 5.v.
[179] Gian Francesco Malfatti, "Memoria sopra un problema sterotomico", *Memorie di Maternatica e di Fisica della Societa Italiana delle Scienze* 10 (1803): 235-244.
[180] H. Lob and H. W. Richmond, "On the Solutions of Malfatti's Problem for a Triangle", *Proceedings of the London Mathematical Society* 30 (1930): 287-304.
[181] Michael Goldberg, "On the Original Malfatti Problem", *Mathematics Magazine*, 40 (1967): 241-247.

- In 1806, André-Marie Ampère claimed to have proved that a continuous function must be differentiable at all but finitely many points.[182] This is untrue: in 1872 Weierstrauss famously gave an example of a nowhere-differential continuous function defined on an interval.[183]

- In 1847, Gabriel Lamé mistakenly thought he had proved Fermat's Last Theorem – most likely following Fermat himself. His mistake was to assume that complex numbers factor uniquely into Gaussian primes.[184]

- A proof of Fermat's Last Theorem was finally found by the heroic efforts of Andrew Wiles and published in the *Annals of Mathematics* in 1995. However, an earlier version of Wiles' argument was found to contain a major error that he spent over a year trying to fix.[185]

- In 1943, Hans Rademacher thought he had disproved the Riemann Hypothesis. An error in his argument was found by Carl Siegel at the last minute, prior to its publication in the *Transactions of the American Mathematical Society*.[186] He erroneously assumed that the logarithm of a complex number was single-valued. The result was nevertheless reported in Time magazine.[187]

- In 1961, an incorrect result regarding Abelian categories that 'many people since have known and used' was published by Jan-Erik Roos. A counterexample was found by Amnon Neeman over forty years later in 2002.[188]

Davis himself recounts a long list of further errors made by mathematicians in print, including the following:[189]

- His own textbook *Interpolation and Approximation* contained some 4 typewritten pages of errata, ranging from 'typos to more serious mathematical errors'.

---

[182] Jesper Lützen, "Between Rigor and Applications: Developments in the Concept of Function in Mathematical Analysis", in *Cambridge History of Science, Volume 5*, ed. Mary Jo Nye, 477.
[183] Karl Weierstrass, "On Continuous Functions of a Real Argument that do not Possess a Well-Defined Derivative for any Value of their Argument", in G. A. Edgar, *Classics on Fractals* (Boston: Addison-Wesley Publishing Company: 1993), 3-9.
[184] Gabriel Lamé, "*Démonstration generale du théorème de Fermat*" in *Compte Rendu des Séances de L'Academie des Science* (1847): 310-315.
[185] Simon Singh, *Fermat's Last Theorem* (London: Harper Perennial, 2002), 277.
[186] Karl Sabbagh, *The Riemann Hypothesis* (New York: Farrar, Strauss and Giroux, 2003), 108-109.
[187] Jonathan Borwein and David Bailey, *Mathematics by Experiment: Plausible Reasoning in the 21st Century* (Massachusetts, A K Peters, 2004), 97.
[188] Amnon Neeman, "A Counterexample to a 1961 "Theorem" in Homological Algebra", *Inventiones Mathematicae* 148 (2002): 397-420. The quotation is from the abstract of this paper.
[189] Davis, "Fidelity in Mathematical Discourse", 261-262.

- The first edition of *A Handbook of Mathematical Functions*, a thousand-page compendium of formulas and tables issued by the National Bureau of Standards, contained hundreds of errors.

- The mimeographed 1925 notes of E. H. Moore on Hermitian matrices was 180 pages long and was appended with 26 pages of errata.

- In 1917, H. W. Turnbull calculated a system of what he thought to be 125 invariants of two quaternary quadratic forms. In 1929, Williamson found that three were reducible; in 1946, Turnbull himself discovered that five more were reducible, and in 1947, J. A. Todd found a further reducible form.

Lastly, Davis also notes that in 1935 a book was published by Maurice Lecat entitled *Erreurs de Mathématiciens des origines á nos jours*, which contained more than 130 pages of errors committed by mathematicians 'of the first and second rank'[190] from antiquity to the turn of the twentieth century: for example, Euler's mistaken assertion that 1,000,009 is prime.[191]

4.v. Public Acceptance and Autonomy

We have seen in the previous sections that under some circumstances mathematicians would be irrational not to yield Private Acceptance to results given by Monte Carlo algorithms. In Section 1.vi, we noticed that some proof presentations are now so long that the checking process is thwarted, so that mistakes become increasingly likely. We also saw that parts of some arguments can only be constructed and checked by a computer. And in Section 4.i, we saw that the iterated structure of IMC testing means that some Monte Carlo algorithms tend to handle errors in a far more robust way than deductive techniques.

Because Monte Carlo strategies are so widely applicable (see Chapter 3), these observations invite the question of whether mathematicians should under some circumstances relax the proof before publication rule and allow Monte Carlo algorithms to feature as justifications in the context of Public Acceptance. Mathematicians such as Phillip Davis have already made similar suggestions:

> 'It is possible that a new type of mathematics might develop in which the "derivations" or the "processes" are so enormously long that the probabilistic nature of the result will be an integral feature of the subject … It is also possible that mathematics might move into a period and into a corpus of material where the proof aspect ceases to have the classical significance and where one can live intimately with less than perfect fidelity.'[192]

---

[190] Davis, "Fidelity in Mathematical Discourse", 262.

[191] Euler himself later showed it was composite by writing it as the sum of two squares in two different ways, and indeed it is equal to 293×3413. He was 70 years old and blind at the time. Leonhard Euler, "An Inquiry Into Whether or Not 1,000,009 is a Prime Number", *Nova Acta Academiae Scientarum Imperialis Petropolitinae* 10 (1797): 63-7.

[192] Davis, "Fidelity in Mathematical Discourse", 260.

Consider also the views of Doron Zeilberger, whose algorithmic identity theory we discussed in the previous chapter. After describing how easily results may be procured with his procedure compared to traditional methods, and regarding the computational cost of solving symbolic equations explicitly, he writes:

> 'As absolute truth becomes more and more expensive, we would sooner or later come to grips with the fact that few non-trivial results could be known with old-fashioned certainty. Most likely we will wind up abandoning the task of keeping track of price altogether, and complete the metamorphosis to non-rigorous mathematics'[193]

For the remainder of the thesis, we will discuss the possibility of revising standards of Public Acceptance, replacing the rule of insistence upon proof discussed in Chapter 1 with the following condition:

ε-condition

A result may be published in a peer-reviewed journal only if there is a proof available, *or* if the result has been produced through a procedure which endorses false results with a probability that is less than $\varepsilon = 10^{-10,000}$.

Here I have deliberately chosen a highly conservative value for the error bound, assuming that all advocates of probabilistic algorithms would find this choice acceptable. As this is also the same error bound that was used with algorithm 3.13, under this revised criterion for Public Acceptance there would be no explicit restriction on our publishing the claim that $66{,}998{,}713$ is prime. Let us now consider what exactly a mathematician announcing the result would need to make public if this were to occur. We first introduce a new definition.

Probabilistic Argument

A discourse giving expression to a chain of reasoning embodied in the arrival at a new result through a Monte Carlo algorithm.

A Probabilistic Argument for a claim is thus analogous to a proof presentation, and can be passed on to others to read with the purposes of convincing them that the conclusion is true. We continue to focus on the Rabin-Miller case. In giving a Probabilistic Argument here, a mathematician might publish the number thought to be prime together with a proof that it meets the ε-condition given above. In addition, they might also include the list of potential witnesses tested. Let us now look at an example of such a discourse. For simplicity, we look for a prime number with only 11 binary digits – that is, an integer N with $1024 \leq N \leq 2047$ – and run only 150 iterations of the test before accepting a candidate. After

[193] Zeilberger, "Theorems for a Price", 7.

running a corresponding program, the number 1729 was produced. We have the following Probabilistic Argument.

Theorem 4.2. 1729 is prime.

Probabilistic Argument: Having run the Rabin-Miller Algorithm to look for primes with 11 binary digits, checking each candidate 150 times before accepting it, 1729 was finally outputted as an answer. The following 150 numbers (arranged here into ascending order) were found as witnesses:

```
9,  10, 12, 16, 69, 74, 75, 81, 90, 92, 103, 108, 120,
129, 144, 160, 166, 172, 173, 181, 191, 192, 235, 256,
257, 263, 289, 302, 347, 355, 363, 365, 374, 376, 386,
402, 426, 433, 438, 439, 443, 484, 493, 536, 545, 555,
562, 563, 568, 575, 584, 621, 625, 638, 649, 653, 654,
666, 675, 690, 699, 706, 729, 740, 750, 757, 797, 802,
807, 809, 810, 828, 829, 831, 841, 857, 872, 888, 898,
901, 919, 920, 922, 927, 932, 972, 979, 989, 1030, 1039,
1054, 1063, 1075, 1076, 1080, 1091, 1104, 1108, 1145,
1154, 1161, 1166, 1167, 1174, 1184, 1193, 1200, 1236,
1245, 1286, 1290, 1291, 1296, 1303, 1327, 1343, 1353,
1355, 1364, 1366, 1374, 1382, 1394, 1427, 1447, 1466,
1472, 1473, 1494, 1537, 1538, 1548, 1556, 1557, 1563,
1569, 1585, 1609, 1621, 1626, 1629, 1637, 1639, 1648,
1655, 1660, 1713, 1717, 1719, 1720
```

It is known that $\pi(2047) = 309$ and $\pi(1024) = 172$, so by Theorem 3.10 the probability that a composite number was selected was less than:

$$\frac{(137/1024)}{4^{150}} \cong 6.568\times10^{-92}$$

Although just a toy example that fails to meet the ε-condition, if a mathematician had run this experiment then they would surely be convinced that the resulting number was prime. The chances of their being so unlucky as to pick 150 strong liars for a composite number are utterly negligible. And the reader can indeed check that none of these numbers are witnesses. However, some readers may feel a creeping sense of suspicion at this point; others may even recall a familiarity with our candidate, perhaps from an anecdote involving Ramanujan.[194] Such suspicion would indeed be warranted: in fact, $1729 = 7\times13\times19$ is not even a prime number at all!

[194] Upon being visited by Hardy in hospital, who asked him if there was any interesting about his taxi-cab number 1729, Ramanujan replied: 'it is a very interesting number; it is the smallest number expressible as the sum of two cubes in two different ways.' $1729 = 9^3 + 10^3 = 1^3 + 12^3$ is sometimes known as the 'Hardy-Ramanujan number'. Godfrey H. Hardy, *Ramanujan* (New York: Cambridge University Press, 1940), 12.

Have we then witnessed a miracle? Not quite: in fact, nothing unlikely has happened at all. Rather than picking 150 numbers at random between 1 and 1728 to test for witnesses, what I really did was simply to test *every* positive integer less than 1729 – which I knew was composite – in order to find all the strong liars.[195] The actual number of strong liars found was 162. I discarded 12 of them at random and deceitfully claimed to have been running 150 tests for each candidate.

Here we see an issue with this particular kind of argument, then. As well as accepting the premises and agreeing with the rules of inference used, the reader will only be convinced by the argument presented if they believe that the author has in fact run the particular experiment they claim to have performed, and is not for example a tricksy young philosopher attempting to pull the wool over their eyes. Probabilistic Arguments could then undermine Autonomy if they became Publicly Accepted because mathematicians reading the argument need to invest trust in their colleagues as well as following their explicit arguments. We discuss this issue further in the next section.

4.vi. Autonomy, Permanence, Reliability and Consensus Revisited

In the previous section, we were concerned that Probabilistic Arguments seem to depend in a central way on the testimony of the author, and that this may damage Autonomy if claims are Publicly Accepted on the basis of these arguments alone.

Discussing this problem, Kenny Easwaran defines an argument as 'transferable' if 'mere consideration of the proposition suffices for a relevant expert to become convinced of the conclusion, unlike arguments in which one needs to know that certain propositions were generated in a suitably random manner, or were generated by a reliable source'.[196]He then points out that proof presentations are transferable. Responding to his article, Don Fallis later agrees that transferability is important for maintaining Autonomy, in something like our sense: 'The transferability of a proof allows a mathematician to check the proof for herself rather than having to rely on the testimony of another mathematician. Thus, the obvious suggestion is that transferability is valuable because it allows an individual to be *epistemically autonomous*.'[197]

However, though we saw in Section 1.vi that having transferable arguments available in the literature is one way of maintaining autonomy, it may not be the only way. Let us now consider what mathematicians might do to remedy their situation, using the above example. Clearly, one cannot in general say anything about whether 1729 was selected randomly from $\{1024, \ldots, 2047\}$, as each number is as likely to be selected as any other. We can however use statistical analysis to

[195] I also suspected that Carmichael numbers such as 1729, for which the consequent of the FLT is true, would tend to have a larger number of strong liars than other composite integers. Of course, there are still less than the global upper bound of $(n-1)/4 = 432$.
[196] Kenny Easwaran, "Probabilistic Proofs and Transferability", *Philosophia Mathematica* 17 (2009): 354.
[197] Don Fallis, "What do Mathematicians Want? Probabilistic Proofs and the Epistemic Goals of Mathematicians" *Logique et Analyse* 45 (2002): 378.

test the hypothesis that the values for witnesses were indeed randomly selected from {1,…,1728}. Let us now do this.

The mean of a discrete random variable distributed uniformly on $\{a, a+1, \ldots, b\}$ is $\mathbb{E}(X) = \mu = \frac{a+b}{2}$ and its variance is $\mathbb{V}ar(X) = \sigma^2 = \frac{(b-a+1)^2-1}{12}$. The Central Limit Theorem tells us that for a random sample $(y_1, \ldots, y_n)$ of size n, with any sampling distribution having mean μ and variance σ^2, the sample mean $\bar{y}$ is approximately normally distributed with mean μ and variance $\frac{\sigma^2}{n}$.

Let H_0 be the hypothesis that our data were selected by a process modeled by a discrete uniform distribution on $\{1, \ldots, 1728\}$. Then the expected value of the sample mean is $1729/2 = 864.5$ with variance $2{,}985{,}583/1800 = 1658.879$. The observed sample mean is 863.047. Performing a two-tailed test, under the null hypothesis the probability of getting a sample whose mean is this much, or even more, less than the true population mean is equal to $\mathbb{P}(Z \leq -0.0357) = 1 - \Phi(0.0357) = 0.486$.

Clearly then there are no grounds to reject the null hypothesis here: the value for the sample mean given by our data seems entirely reasonable. We could also test the sample variance, or perform another test such as the χ^2 test. However, no such test is ever completely conclusive, and it does look very much like our data were generated by a uniformly distributed random variable – although we know that this is in fact untrue.

A more successful strategy for maintaining Autonomy is for readers to generate their own potential witnesses.[198] Suppose that we read a Probabilistic Argument for the claim that some number N is prime. We then randomly select 50 numbers from $\{1,2, \ldots, N-1\}$, none of which are witnesses. How confident should we now be that N is indeed prime? Again, to answer this question we need the prior probability of N's being prime. But because we did not select N ourselves and were merely presented with it, there is no acceptable way of assigning a determinate value here, for reasons made plain above. And so we cannot acquire direct mathematical reasons that enable us to Autonomously assign a determinate probability to the belief they support – that N is prime – having been false.

However, as we have tested a large number potential witnesses, if N turns out to be composite we would clearly have observed a miracle. So it is reasonable to appeal to an inference to the best explanation here: on the basis of our testing, we can conclude that the work was not fabricated. But if we do come to accept that the publishing mathematician has performed the Monte Carlo procedure correctly, then we will also have excellent reasons for believing that N is indeed prime.

Though the reader's ultimate reasons for believing N is prime are not expressed within the published discourse itself, Autonomy can nevertheless be maintained. For the modal condition that any competent researcher can come to have their own direct reasons for believing any Publicly Accepted claim is still met, as there is a

[198] As pointed out by Fallis. Ibid., 381.

clear procedure for generating witnesses that will always work if the claim is true. If it does not work, and a witness is found, the claim must be retracted – much like finding an unbridgeable gap in a proof presentation. The second condition for Autonomy can be met, too. Mathematicians need never be permitted to publish a result on the basis of trust or authority, as journal referees can also generate their own potential witnesses to check the work has not been fabricated.

We now move on to a discussion of Permanence and Reliability. Whilst the discussion of human error in the previous section was rather anecdotal in flavour, it is nevertheless clear that replacing the requirement of proof with the ε-condition is not a serious reason for concerns about either of these two Practical Virtues. In Chapter 1 we noted that an estimated 200,000 proof presentations are published each year, and mistakes can and have been found. Yet if Probabilistic Arguments meeting the ε-condition were accepted, the overwhelming probability is that not a single Implementation Error would ever occur in the entire future of the universe, even if these algorithms become used vastly more frequently than deductive algorithms are currently used in mathematics today.

Lastly, we briefly discuss Consensus. It is plainly possible that some mathematicians would simply refuse to believe results whose only justification was given by a probabilistic algorithm. Yet it is not merely the sociological fact of Consensus that we are taking to be valuable here – the brute fact that there is *de facto* agreement – but rather that the mathematical community has the intellectual resources to provide its members with rationally compelling reasons to believe the truth of new results. And the arguments of this chapter show that if mathematicians are not to be convicted of irrationality then they must come to believe results justified by randomised algorithms meeting the ε-condition. There can therefore be no new rational disagreement introduced into the context of Public Acceptance by this revision to established practice.

4.vii. Conclusion

In this chapter, I have argued that to maintain agnosticism about results delivered by Monte Carlo algorithms whilst yielding credence to deductive algorithms for the same task that are known to be less reliable overall would be a failure of epistemic rationality. Furthermore, due to their robust iterative structure, Monte Carlo algorithms tend to exhibit superior reliability in practice.

We then considered the possibility of mathematicians replacing their rule of proof prior to publication with a weaker ε-condition: that only results for which *either* a proof is available *or* which have been endorsed by a procedure that endorses false conjectures with probability less than $\varepsilon = 10^{-10,000}$ are eligible for Public Acceptance.

In the final section, I argued that this revision to established practice would not diminish the extent to which mathematics embodies our four Practical Virtues. If these Monte Carlo methods are available for the kinds of problems discussed in Section 1.vi, where proof can supply only a less effective means of justification, perhaps they can even help to promote them.

5. Normative Standards Within Mathematics

In the opening section of this chapter, we examine an existing approach to the question of whether mathematicians should replace their restriction of proof prior to publication with something like the ε-condition. This is the means-ends reasoning framework of philosopher Don Fallis. In the second section, I argue for an alternative way of looking at the problem, wherein we regard mathematical discourses as governed by shared, normative standards that emerge historically. The remainder of the chapter will describe four such norms pertaining to contemporary mathematics, together with historical illustrations and reasons as to why they are now important for its continuing success.

5.i. Means-Ends Reasoning and the Epistemic Objectives of Mathematicians

In a 2002 paper, philosopher Don Fallis discusses mathematicians' ongoing rejection of Probabilistic Arguments and examines potential reasons for this feature of mathematical practice.[199] By 'rejection' he means that mathematicians never take these arguments to 'establish' a result, which we can take to mean something like regarding them as providing an argument sufficient to warrant Public Acceptance. As the focus of his paper is also the Rabin-Miller Algorithm, we will examine his argument in this section.

Fallis attempts to explain mathematicians' rejection of Monte Carlo methods by employing a means-ends reasoning framework. He first suggests various 'epistemic goals' that individual mathematicians share, and then enquires as to whether their having these particular goals suffices to explain their rejection of Probabilistic Arguments. He identifies three reasons why mathematicians might choose to avoid using Probabilistic Arguments in light of their goals. Firstly, they might recognise that Probabilistic Arguments are not suitable means for achieving these goals. Secondly, it might be the case that although these arguments are in fact suitable means to their goals, this is not realised by mathematicians themselves. Thirdly, mathematicians might not be rational. He further restricts his enquiry to seeking explanations of the first type.[200]

The focus of Fallis' paper is also on specifically epistemic[201] goals, though he is well aware that there are other goals mathematicians could have. He takes the central epistemic goals of mathematicians to be acquiring more true beliefs and avoiding false ones, but considers other, more specific epistemic goals as well. Before we move on to his argument, we first consider some other goals that

[199] Don Fallis, "What do Mathematicians Want? Probabilistic Proofs and the Epistemic Goals of Mathematicians."
[200] Ibid., 6.
[201] Though Fallis does not give a clear indication of the sense in which 'epistemic' is intended, it is meant in a fairly broad sense to include for example understanding as an epistemic goal.

mathematicians have that proofs can be an effective means to achieving, some of which are not epistemic in nature.

Firstly, the search for a proof of a theorem can enable us to see connections between different areas of mathematics, and even suggest a new theorem, such as a generalisation. This idea of connectedness is central to Hardy's concept of what it is to do good mathematics:

> 'The 'seriousness' of a mathematical theorem lies, not in its practical consequences, which are usually negligible, but in the *significance* of the mathematical ideas which it connects. We may say, roughly that a mathematical idea is 'significant' if it can be connected, in a natural and illuminating way, with a large complex of other mathematical ideas.'[202]

To take a concrete example, following the work of Gerhard Frey and Jean-Pierre Serre, Ken Ribet showed that Fermat's Last Theorem would follow from the Taniyama-Shimura Conjecture. This paved the way for Andrew Wiles' proof of the theorem and connected number theory with the study of modular forms within topology.[203] De Villiers summarises the point more generally:

> 'Clearly, the value here is largely in gaining multiple perspectives; developing a deeper, richer understanding; or opening up a whole range of possible analogies, connections, specializations, and generalizations that can be further explored.'[204]

The challenge of finding a proof is also a spur to developing new mathematical tools: to prove the theorem Wiles had to collect and refine an impressive array of techniques in number theory. The hunt for a proof can also help us solidify our mathematical concepts, as what is really essential to them must be clarified to give a rigorous deductive argument. For instance, in Lakatos' *Proofs and Refutations*, he describes successive and increasingly adequate attempts at defining polyhedra when proving Euler's Theorem.[205]

Next, within an educational context, seeing the logical connections between concepts is also a useful means to acquiring and fully understanding them. Hence, the focus in pure mathematics lecture courses is typically on definitions, theorems and proofs, together with a few examples. At a more elementary level, perhaps no mathematician has done more in recent years to highlight the benefits of introducing proof into the classroom than De Villiers.

Lastly, proofs can also provide aesthetic pleasure. This may have more impact on mathematicians' research than we might initially think. Assuring us this attitude is widespread and the norm, Hardy poignantly writes:

---

[202] Godfrey H. Hardy, *A Mathematician's Apology* (Cambridge: Cambridge University Press, 1992), 89.
[203] Ken Ribet, "On Modular Representations of $Gal(\bar{Q}/Q)$ Arising From Modular Forms" *Inventiones Mathematicae* 100 (1990): 431-476.
[204] De Villiers, "The Role and Function of Quasi-Empirical Methods in Mathematics", 413.
[205] Imre Lakatos, *Proofs and Refutations* (Cambridge: Cambridge University Press, 1976).

> 'Beauty is the first test; there is no permanent place in the world for ugly mathematics.'[206]

Valuable as these benefits of proof are, they will not help us to explain mathematicians' ongoing rejection of Probabilistic Arguments. For clearly, they are features belonging only to some but not all acceptable deductive arguments. If a proof of an important theorem is found, then journals will go ahead with publication regardless of whether the proof affords deeper insight into connections between results, can be generalised, or is aesthetically pleasing. Hence, an argument's being a means to securing one of these ends is desirable but not essential within the context of Public Acceptance. Consider in particular the experiments of chapter 3: as Fallis later says, we must provide an explanation that 'is consistent with the acceptability of the trial division test', and Algorithm 3.12 clearly has no such benefits.[207]

We now move on to the rest of Fallis' argument. The epistemic objectives he imputes to mathematicians fall into two categories: those that do not explain the rejection of Probabilistic Arguments, and those that offer only a partial but unsatisfying explanation. We consider each possible objective in turn.

Epistemic objectives that do not explain

1. Epistemic Conservativeness

As mentioned above, Fallis begins from the premiss that mathematicians seek to acquire true mathematical beliefs whilst avoiding believing falsehoods. But compared to other scientists, who are willing to use inductive methods, mathematicians are especially 'epistemically conservative'. That is, they put a greater premium on avoiding error, even at the cost of greatly slowing down the rate at which new true mathematical beliefs are accumulated. However, Fallis points out that imputing this goal to mathematicians does nothing to explain the rejection of Probabilistic Arguments, since they are willing to make use of deductive methods that are in practice less reliable overall.

2. Long Term Errors

The second epistemic objective he considers is that mathematicians want to avoid errors 'in the long run', which seems to mean that any false beliefs they do acquire are found as swiftly as possible. Though the first draft of a proof presentation may contain errors, most of the time these will be uncovered by the reviewing process. However, mathematicians can also catch errors made through the use of Monte Carlo methods, by running more iterations of the testing phase themselves. Probabilistic Arguments can therefore also be an equally adequate means of avoiding errors in the long term.

---

[206] Hardy, *A Mathematician's Apology*, 85

[207] Fallis, "What do Mathematicians Want?", 13. See also Don Fallis, "The Epistemic Status of Probabilistic Proof", *The Journal of Philosophy* 94 (1997): 165-186.

3. Developing Consequences of Results

We have seen that mathematics is built up like a house of cards, with arguments for new results usually relying on other results taken to have been established previously. So rather than considering results one by one, mathematicians might then have the epistemic goal of avoiding errors creeping into this structure, to prevent the entire thing from coming tumbling down. But again, where probabilistic methods are more reliable than deductive ones, they are also more suitable as a means to keep the whole mathematical edifice free from errors too.

4. Understanding

Lastly, Fallis lists understanding as another epistemic objective: it has long been acknowledged by mathematicians and philosophers that proofs can enable us to see *why* a result is true as well as *that* a result is true. The point is well summarised by Yu Manin: 'A good proof is a proof that makes us wiser.'[208] However, it is unclear what sense may be attached to the question of *why* 66,998,713 is prime. And again, Fallis points out the Trial Division Algorithm will also not provide any understanding of why the result is true. Both algorithms from the last chapter gave us no information other than the single integers they outputted.

Epistemic Objectives that do explain but that are not satisfying

1. Proofs as Intrinsically Valuable

Next, Fallis considers that deductive arguments might be of value for their own sake, rather than as means to an end; that the construction of proofs could be a legitimate goal in its own right. He is rather dismissive of this idea, writing: 'We might have hoped that mathematicians restrict themselves to using deductive proofs because doing so is the most effective means to achieving some further epistemic objective (such as avoiding errors and finding errors that have already been made).'[209]

2. Liability for Mistakes

Another potential epistemic objective is that mathematicians might want to avoid coming to believe falsehoods simply by being unlucky and through no fault of their own. If a mathematician comes to believe a false claim for which they think they have a deductive proof then they must have made a mistake somewhere in their argument, but a Monte Carlo algorithm might yield an incorrect result even if it has been implemented perfectly. However, it is mysterious exactly why this would be of value to mathematicians, and in any case mathematicians accept results yielded by deductive methods that are subject to various unpredictable sources of error, as discussed earlier. They could therefore be unlucky here too. Using Monte Carlo algorithms instead also reduces the chances of such errors.

[208] Yuri I. Manin, "Good Proofs are Proofs that Make us Wiser", interview by Martin Aigner and Vasco A. Schmidt, *The Berlin Intelligencer* (1998): 16-19.
[209] Fallis, What do Mathematicians Want?", 16.

3. Correctness in Principle

A third, related objective mathematicians might have is aiming to use only techniques that always work correctly in principle, if they are applied properly and put into practice without incident. Again, we saw in the last chapter that this consideration also does not provide convincing epistemic reasons for the rejection of probabilistic methods. We know that we live in a world where such implementations errors do occur, and furthermore 'we would have hoped that mathematicians use deductive proofs exclusively because of some *actual* epistemic benefit that they derive.'[210]

4. Applicability

The last epistemic goal Fallis considers is that mathematicians might prefer techniques that are widely applicable. Deductive techniques can be used in a wide variety of situations, whereas Probabilistic Arguments may currently be somewhat more limited. However, this does not explain the rejection of Probabilistic Arguments where they are available and known to be reliable. This is especially puzzling in cases where randomised algorithms are the only available way of solving a problem that is too complex to be amenable to deductive techniques. Moreover, we saw in Chapter 3 that randomised algorithms can already be employed in quite a broad range of situations, even though the field is perhaps still in its infancy. In any case, they are not intended to replace deductive methods entirely, but merely to constitute one more tool in a mathematician's arsenal.

Having considered all the epistemic goals available in the literature, and found none that might explain mathematicians' rejection of Probabilistic Arguments, Fallis concludes that the burden is now upon mathematicians to come up with goals that make sense of their rejection. If they are unable to do so then presumably we are to conclude either that mathematicians have yet to realise that Probabilistic Arguments are means to their goals, or that they are behaving irrationally in continuing with this aspect of mathematical practice.

5.ii. The Rationality of Public Acceptance

Let us take it that Fallis' arguments are correct on his own terms. In fact, it seems reasonable to draw even stronger conclusions from the above considerations than he does. Within his individualist means-ends framework there are no epistemic reasons for mathematicians to prefer deductive algorithms over Monte Carlo procedures which are known to be more reliable overall. Nevertheless, I shall argue for the conservative view: that mathematicians are rational in continuing to reject these Probabilistic Arguments in the context of Public Acceptance. For like the infallibilist in Section 4.ii, we shall see that Fallis has been misled by taking too narrow and restricted a view of how the problem is to be approached.

[210] Ibid., 16.

My response to Fallis' argument has four parts. Firstly, I reconsider the relationship between mathematicians' individual goals and the overall progress of the discipline. Secondly, I argue that as well as being determined only with reference to individual goals, the rationality of the Public Acceptance of new results can also be understood in terms of adherence both to shared rules and to shared standards of excellence. Thirdly, we see that these rules and standards of excellence are general and extend across all of mathematics, rather than being highly particularised and judged according to individual cases. Fourthly, the standards of excellence pertain to finished mathematical discourses.

Towards the end of his paper, Fallis defends the individualist framework he has adopted, wherein he takes as his starting point the *de facto* personal goals of individual mathematicians and regards the content of rationality as pertaining to the selection of appropriate means to further the pursuit of these goals.[211] He concludes that to adopt a strategy that does not have the best individual epistemic consequences overall seems rather perverse. This is a view that fits well with the line of argument given in the previous chapter, and his framework does accurately capture the situation of a mathematician deciding whether to Privately Accept a conjecture. However, it is somewhat misleading when we consider the shared standpoint of Public Acceptance (Fallis himself does not discuss the Public-Private partnership explicitly).

It is true enough, of course, that mathematicians may have any number of individual goals, motivations and desires that drive the direction of their research. In Hardy's famous memoir, he writes:

> 'There are many highly respectable motives which may lead men to prosecute research, but three which are much more important than the rest. The first (without which the rest must come to nothing) is intellectual curiosity, desire to know the truth. Then, professional pride, anxiety to be satisfied with one's performance, the shame that overcomes any self-respecting craftsman when his work is unworthy of his talent. Finally, ambition, desire for reputation, and the position, even the power or the money, which it brings.'

Yet equally clearly, when a mathematical discourse is to be submitted for publication, its author(s) must contend with the publication rules and shared standards of excellence held in place by both journals and by the mathematical community itself. Mathematics is an example of a practice, in the sense given to this term by Alasdair Macintyre.[212] The standards of excellence characterise what it is to do mathematics well, and must be learnt by its newest practitioners if they are to contribute to research within the field.

---

[211] In a later paper, Fallis also considers 'shared goals' of mathematicians – but from the perspective of the objections given here his position is not relevantly different. Don Fallis, "Probabilistic Proofs and the Collective Epistemic Goals of Mathematicians", in *Collective Epistemology*, eds. Hans Bernard Schmid, Marcel Weber, and Daniel Sirtes (Germany: Ontos Verlag, 2011), 157-175.

[212] Alasdair Macintyre, *After Virtue*, (London: Bloomsbury, 2011), 218.

It is not just that personal motivations are subordinate to these shared rules and standards, and that mature mathematicians' individual goals can only be pursued through submitting to them. As we seek to progress in our mathematical education, at first we are not primarily attempting to satisfy our own predetermined epistemic goals at all. Rather, we are learning to think like a mathematician:

> 'To enter into a practice is to accept the authority of those standards and the inadequacy of my own performance as judged by them. It is to subject my own attitudes, choices, preferences and tastes to the standards which currently and partially define the practice.'[213]

From the perspective of the practice then, what is or is not a good reason for using a certain technique or adopting a certain approach to a problem – or even for whether a certain problem is a good one – is encountered as something that is independent of and prior to my initial current goals and choices. Coming to understand the distinctive form of giving and asking for reasons amongst mathematicians in the context of Public Acceptance may require a long and arduous training, and is not available simply through a general enquiry into the logic of epistemic vocabulary. It is only by becoming initiated into the practice, and learning through the example of others how to reason like a mathematician and to produce good mathematics, that our mature research goals as developed mathematicians can later be formulated.[214]

Understanding and evaluating the form of giving and asking for reasons for the Public Acceptance of new results that is embedded within mathematical practice requires a pluralistic approach. Firstly, we have seen that mathematical research is structured by shared rules. For instance, mathematicians require proofs for the Public Acceptance of new results. Of course, we are not able to invoke this particular rule here, as precisely what is in question is its rational justification: whether continuing to uphold it is really in mathematicians' best interests, or is contributing to the flourishing of mathematics itself. But notice that it extends across the discipline as a whole, rather than having its applicability decided upon in a piecemeal way, by consideration of individual cases.

Secondly, mathematics is also characterised by the four Practical Virtues that are important for mathematics to continue to flourish and for mathematical enquiry to achieve the high level of success it has enjoyed hitherto. These concern the body of accepted mathematics itself (Reliability), the relation of the community of researchers to this mathematical edifice over time (Permanence), the relationship of researchers to each other (Consensus), and the relationship of individual researchers to the community as a whole (Autonomy).

---

[213] Alasdair Macintyre, *After Virtue*, 221.

[214] There are of course exceptional individuals such as Ramanujan who learn to do valuable mathematics outside of the mainstream tradition. Such individuals will often create their own way of doing things: for instance, Ramanujan notoriously did not share our concept of proof or view of its importance. But these are the exception rather than the rule, and their uniqueness only highlights the rule by being so widely remarked upon.

The Practical Virtues may not have as much weight in determining mathematicians' choices in comparison with their individual goals. But it is clear that we can and should evaluate mathematicians' collective choices in terms of the maintenance of the Practical Virtues, and that this project is clearly a very different project from the one Fallis pursues. Yet I have also argued that the Public Acceptance of Probabilistic Arguments would not undermine the Practical Virtues. It seems that we can envision an equally robust research community where practitioners accept Probabilistic Arguments as binding in the context of Public Acceptance, and mathematical enquiry continues to flourish.

However, I have also indicated that a third perspective is available here: the production of new mathematics is also regulated by shared standards of excellence. Our quartet of Practical Virtues will thus be complemented by a suite of what I will call 'Intellectual Virtues'. These are four normative standards that apply directly to published mathematical discourses themselves: ideals that for a time supply us with constraints upon what a good piece of mathematics can be.[215]

These Intellectual Virtues can and do influence the individual choices of mathematicians because they bear directly on what kinds of reasoning are acceptable to express in published discourses in order to establish a result as Publically Accepted. Mathematics is therefore less exclusively results-orientated than Fallis takes it to be. Mathematicians do not simply pick up any means that are suitable for the justification of their results according only to a generalised concept of individual epistemic rationality. Rather, each concrete mathematical achievement considered as a whole must adhere to these shared standards of excellence.[216] Our four Intellectual Virtues are given as follows:

Abstractness

Mathematics concerns *abstracta* and never makes claims essentially referring to spatiotemporal particulars.

Explicitness

Published arguments are always capable of being made explicit and never require complex intuitive leaps that are not reducible to simple steps.

Univocality

Concepts essentially used in published mathematical arguments (other than perhaps a few basic concepts) are always attended with precise necessary and sufficient defining conditions, and any specific entity essentially referred to is always given a clear definite description.

[215] There is some overlap here with Section 1.iv, where we discussed adequacy conditions for proof presentations, though the focus is now different.

[216] For further criticism of the view of mathematics as primarily results-oriented, see David Corfield, *Towards a Philosophy of Real Mathematics* (Cambridge: Cambridge University Press, 2004), 181.

Formalizability

Acceptable mathematical arguments are always in principle capable of being set out within a standard formal system, for which the axioms and rules of inference may be given fully explicitly.

Examination of these standards will also reveal that although they have a broad application across the discipline as a whole they are neither eternal nor immutable. Each has a history and has emerged only through the continuing tradition of mathematical enquiry up to this point – often in response to highly specific problems. In explaining them further, then, I will make use of historical illustration from periods that I take to be especially crucial in their development.

For Abstractness, the chief illustration will be the conception of mathematics as concerning propositions about abstract ideal spatial objects found in some strands of Ancient Greek geometry, such as the research carried out at Plato's Academy. To understand Explicitness, we discuss the replacement of visual and intuitive techniques with more systematic algebraic and symbolic methods attendant to the development of analytic geometry in the 17th Century, and the parallel break with intuition initiated within synthetic geometry. Our chief historical example for Univocality is the clear formulations of limits, functions and derivatives arrived at by Cauchy and Weierstrass in the search for foundations for the calculus in the 19th Century. And naturally we discuss Formalizability in the context of the deeper foundational research carried out in the 20th Century, motivated by the goal of achieving for mathematical proofs the ideal transparency of formal logic. We can also see that these developments are linked and that there is progress towards a distinctive kind of excellence that is characteristic of mathematics as such.

We will however need to go further than merely writing history here. These standards of excellence must themselves be susceptible to clear rational justification if the practice of mathematics is not to become arbitrary and dogmatic. It will be insufficient only to identify the historical development of the Intellectual Virtues, or even the ways in which their adoption was necessary to overcome specific problems that mathematicians faced at the time. For it may be that they have since become obsolete, and their observance is no longer necessary.

How is this rational justification to be achieved? Clearly, Fallis' program could resurface at this point, as it may be possible to justify the pursuit of the Intellectual Virtues with reference to mathematicians' goals. But a more primary question is whether observance of the standards that these virtues embody is contributing or detracting from the flourishing of mathematical enquiry: whether or not they are essential for the practice of mathematics as it now exists to continue in good working order. As we shall see in this chapter, these four Intellectual Virtues do facilitate the development of mathematical research, for two kinds of reasons: both directly and because they maintain and enhance the Practical Virtues.[217]

In arguing for the importance of these four Intellectual Virtues in the remainder of this chapter, we will also prepare the way for the final argument against Monte

[217] It is arguable that Reliability should be grouped with the Intellectual Virtues.

Carlo methods, to be given in the next chapter. Here we will show that Probabilistic Arguments inevitably fail to comply with these standards of excellence. This provides mathematicians with strong reasons for rejecting them in the context of Public Acceptance.

5.iii. Mathematics as Concerning Abstracta

> 'A mathematician, like a painter or a poet, is a maker of patterns. If his patterns are more permanent than theirs, it is because they are made with *ideas*.' – *G. H. Hardy*[218]

In this section, I shall describe how the propositions given in a piece of mathematics never make essential reference to particulars situated in space and time, and always concern or can always be interpreted so as only to concern abstract entities. We begin with some historical illustration.

Sometime around the start of the 5th Century BCE, the science of pure geometry was established by the Ionian Ancient Greeks. Though we now possess only fragments and scattered references from later authors, it is clear that during this period discourses concerning abstract planar and spatial entities were produced. By the time of Euclid's Elements – conventionally dated to the year 300 BCE – this research programme had crystalised into a unified and elaborate deductive axiomatic system in which theorems are derived from a small number of widely accepted premisses.[219] Here for the first time we find a general and persistent search for stable universal truths underlying the chaos and flux of physical, empirical reality; a tendency that permeates Ancient Greek natural philosophy more generally. In contrast, according to Pascal Boyer, pre-Hellenic peoples:

> 'lacked the tendency, essential to both mathematical and scientific method, toward the isolation of certain samenesses from their confusingly carried concomitants in nature and in thought. Lacking these elements of invariance to serve as premises of inference, they were accordingly without appreciation of the characteristics which distinguish mathematics from science, namely, its logical nature and the necessity of proof.'

As made clear by his dialogue *Republic*, Plato regarded pure mathematical geometry – as distinct from practical geometry and mensuration – as having great intellectual value, and he encouraged its study at the Academy.[220] Though Plato agreed that the physical world was to a large extent imperfect, unstable and unknowable, the geometricians at the Academy were not concerned with empirical phenomena at all, but rather with a perfect, abstract and unchanging world beyond

---

[218] Hardy, *A Mathematician's Apology*, 84.

[219] See Wilbur Richard Knorr, *The Ancient Tradition of Geometric Problems* (New York: Dover, 1993); Euclid, *The Thirteen Books of the Elements, 3 Volumes*, trans. with introduction and commentary by Thomas Heath, (New York: Dover, 2012); and Proclus, *A Commentary on the First Book of Euclid's Elements*, trans. with an introduction by Glenn Morrow (Princeton: Princeton University Press, 1992).

[220] David Fowler, *The Mathematics of Plato's Academy: A New Reconstruction* (Oxford: The Clarendon Press, 1991).

physical reality. This kind of mathematics was a good example of human beings attaining unqualified knowledge of absolute, objective truths. In *Republic*, Plato emphasises that although mathematicians use visible diagrams in their thinking, these are not the true objects of their enquiry. Socrates explains to Glaucon:

> 'And you will also be aware that they summon up the assistance of visible forms, and refer their discussion to them, although they're not thinking about these, but about the things these are images of. So their reasoning has in view the square itself, and the diagonal itself, not the diagonal they have drawn. And the same with other examples.'[221]

Though it is perhaps arguable that in the Renaissance mathematics was seen as about the corporeal world, the Abstractness of this strand of Ancient geometry is now characteristic of contemporary mathematics more generally.

We now pause here to consider a potential rejoinder to this claim. For even in looking back through the small number of mathematical problems that have been discussed in this thesis so far, such as the combinatorial questions from Chapter 1, or the problem with Alice and Josh jumping down the stairs in Chapter 2, we find both questions and solutions apparently referring to an array of physical situations involving committees, tennis tournaments, staircases, and necklaces.

These references are only for psychological convenience, however, as the final proof of Fermat's Little Theorem illustrates. The other arguments can also be recast in a similar form, and so do not constitute counterexamples to the Abstractness of contemporary mathematics. Moreover, the kind of knowledge we gained here was clearly different in character from knowledge about the physical properties of the phenomena referred to, and surer and more certain than such knowledge could ever be. What is the average weight of a necklace? What proportion of tennis tournaments is won by a winner from the previous year? Consider again another concrete example.

Problem 5.1.

> Suppose we have a bench one metre long, and an unlimited supply of ants, which walk at one metre per minute. We place the ants anywhere on a thin line running along the bench, facing either direction, after which they will begin walking forwards. If two ants bump into each other, they will collide perfectly elastically: that is, each will instantaneously turn through 180 degrees on their axis and continue in the other direction (the ants may be considered to be of negligible size). How should we place the ants on the bench so that at least one ant stays on the bench as long as possible?

The solution may be reached as follows. Considering the local interactions between ants, we notice that for a given arrangement of ants, they will stay on the bench exactly as long if we change the rules so that instead of bumping into each

[221] Plato, *The Republic*, trans. Tom Griffith, ed. G. R. F. Ferrari (Cambridge: Cambridge University Press, 2007) 217-218 (510c,d).

other and turning instantaneously they simply walk through each other without influencing each other at all. We can now easily see that the maximum time is one minute, which will always be attained so long as at least one ant is placed at either end of the bench faced toward the other end.

Now, superficially, both problem and solution make reference to ants and to corporeal concepts such as 'bumping' and perfectly elastic collisions. But one can see this is easily avoidable; we can see and feel the precision in the underlying mathematical argument, though it might require a bit of thought to recast it in a more acceptable guise. The reference to the physical world is thus not essential, and again there are no grounds for objection to the core claim of this section. And the same is true in similar cases.

We now say a few words about why Abstractness is of ongoing importance for mathematics today. Firstly, as we have already claimed, the certainty of mathematics is much higher in comparison with even the most fundamental items of physical knowledge, and this increased certainly in comparison with natural science is surely in part due to the ideal character of mathematics. For instance, mathematicians never make claims of the form '*X* is a group', where *X* is a collection of physical entities or patterns of interactions within a physical system. They only discuss what would follow *if* the components of a system did form a group by drawing out the logical consequences of the axioms. Mathematicians therefore avoid committing themselves to empirical assumptions that are unnecessary for their peculiar work. Yet within physics it is precisely these empirical modeling assumptions that cannot be conclusively established:

> 'The certainty of mathematics depends upon its complete abstract generality. But we can have no *a priori* certainty that we are right in believing that the observed entities in the concrete universe form a particular instance of what falls under our general reasoning.'[222]

In addition, the high level of abstraction we find in mathematics leads to an impressive economy of thought. Mathematical techniques can be applied to give accurate results across a wide variety of practical problems, resulting in what Wigner has called the 'unreasonable effectiveness' of mathematics in natural science.[223] For instance, distinct abstract geometrical concepts such as 'straight line', 'circle', 'right-angle' can be applied to problems in many diverse fields in order to aid the thinking of a wide variety of practitioners.[224]

Furthermore, because its content is abstract, mathematics is not restricted to concepts that have sensory counterparts. It now seems clear that we are 'at liberty arbitrarily to create imaginaries',[225] even if these correspond to nothing in nature.

---

[222] Alfred North Whitehead, "Mathematics as an Element in the History of Thought", in *The World of Mathematics*, *Vol 1*, ed. James Newman (New York: Dover, 1956) 404.
[223] Eugene Wigner, "The Unreasonable Effectiveness of Mathematics in the Natural Sciences," *Communications on Pure and Applied Mathematics* 13 (1960): 1-14.
[224] On this point see Philip Jourdain, "The Nature of Mathematics", in *The World of Mathematics*, *Vol 1*, ed. James Newman (New York: Dover, 1956), 14-15.
[225] John Graves' famous reservation about Hamilton's definition of quaternions. Quoted in Michael Crowe, *A History of Vector Analysis* (Toronto: University of Notre Dame Press, 1967), 34.

No mathematician today is worried about whether quaternions or complex numbers 'really exist', in the purely metaphysical sense – they are simply identified with certain sets of real numbers subject to formal algebraic rules. This freedom to go beyond nature enables the development of systems that are far more harmonious and complete, which ultimately leads to advances in applications. Computation of real integrals is often much easier using complex analysis, for instance, and the use of complex numbers is common in engineering.[226]

5.iv. The Decline of Visual Intuition

In this section I claim that in contemporary mathematics, Publicly Accepted Results are always attended by arguments that it is possible to reduce to clear and simple steps, and never rest only upon leaps of intuition that cannot be broken down into such steps. Our story picks up in the 17th Century, during which time direct reasoning about spatial entities in geometry (as in the arguments of Euclid's Elements, Books I and II) comes to be supplemented with Descartes and Fermat's algebraic-symbolic methods, made possible by the use of equations to define curves. These new techniques, which later come to predominate, permit more explicit inferences and proofs that do not rely on visual intuition. Moreover, synthetic geometry – which continues to reason directly about geometrical entities – also develops to become increasingly abstract, culminating in Hilbert's axiomatic approach, which is again entirely divorced from spatial intuitions.

In his 1637 work *La Geometrié*, Descartes showed how arithmetic operations such as extracting a root can be performed geometrically. He also advocated the view that if a represents a line segment, then a^2 should not represent an area, but rather a second line segment that stands in the same proportional relationship to a as a does to a line segment of unit length:[227] that is, $1 : a = a : a^2$. Once all such symbolic expressions are so interpreted, then 'Any problem in geometry can be reduced to such terms that a knowledge of the lengths of certain straight line segments is sufficient for its construction.'[228] Descartes also came to identify curves in the plane given by loci with algebraic equations in two variables. In the words of Fermat:

> 'Whenever in a final equation two unknown quantities are found, we have a locus, the extremity of one of these describing a line, straight or curved.'[229]

Descartes and Fermat had in fact already worked out the basic principles of analytic geometry by the 1620's, having realised that 'all the properties of a curve such as the magnitude of its area or the direction of its tangent are fully determined when an equation in two unknowns is given'.[230] Descartes also announced a general method of finding the normal to a curve at an arbitrary point, by considering the radius of a circle touching the curve. In hindsight this seems

[226] Though some engineers have been known to have trouble spelling them.
[227] René Descartes, *La Geometrié*, in *God Created The Integers,* ed. Stephen Hawking (London: Penguin, 2004), 293.
[228] Descartes, *La Geometrié*, 292.
[229] Quoted in Pascal Boyer, *History of Analytic Geometry* (New York: Dover, 2004), 75.
[230] Uta Merzbach and Carl Boyer, *A History of Mathematics*, 318.

highly promising, though his technique is hard to apply in practice, and like many of his other ideas was not worked out fully in his published writings.

In 1659 and 1661, van Schooten published a large two-volume second edition of his *Geometria a Renato Des Cartes*, with elaborate commentary and explanation to make up for the rather obscure style of Descartes' 1637 work. This edition included treatises by DeWitte giving the explicit equations for conic sections, techniques by Sleuse for finding the tangent to any polynomial curve and a variety of others, methods by Huygens and Hudde for finding points of inflexion and maxima and minima, and much else besides.[231] The work was highly influential and was read by Newton and Leibniz, facilitating their later work on the general algebraic study of curves.

We pause here to consider a simple example of the power of analytic geometry, using modern notation and what are now called 'Cartesian co-ordinates'.[232]

Problem 5.2.

Let ABC be an equilateral triangle. Let D be on AB produced such that $AB = BD$. Let E be on BC produced such that $BC = CE$. Let F be on BC between B and C such that $2BF = FC$. Let X be the intersection of the lines through AF and DE. Find the angle $\angle BXD$.[233]

This problem may require some thought if approached in a traditional geometric manner, reasoning directly about the geometric objects defined. However, using techniques from co-ordinate geometry in its modern form the solution will be entirely mechanical.

We first interpret the triangle as a certain subset of points on the Cartesian plane. Without loss of generality, let the vertices A, B, C be at $(0,0)$, $(1,0)$ and $(1/2, \sqrt{3}/2)$ respectively. Then D is at $(2,0)$, E is at $(0, \sqrt{3})$, and so F is at $(5/6, \sqrt{3}/6)$. Hence the line through $A(0,0)$ and $F(5/6, \sqrt{3}/6)$ is given by $y = \sqrt{3}/5\, x$ and the line through $D(2,0)$ and $E(0, \sqrt{3})$ is given by $y = \sqrt{3} - \sqrt{3}/2\, x$. Solving these two equations together, we find that X is at $(10/7, 2\sqrt{3}/7)$. It follows that the gradient of BX produced is $2/\sqrt{3}$, whereas the gradient of DX produced is $-\sqrt{3}/2$, meaning that $\angle BXD$ is a right angle.

Once the power of these techniques was appreciated, there was a recognisable move towards an emphasis on algebraic rather than geometric derivation in the

[231] Jan van Maanen, "Precursors of Differentiation and Integration", in *A History of Analysis,* ed. Hans Niels Jahnke (London: London Mathematical Society, 2003),47.

[232] The development of co-ordinate geometry has a long history. Ancient geometers such as Apollonius and Archimedes used auxiliary lines in their constructions, and Nicole Oresme (1323-1382) represented functions graphically in a co-ordinate system. Co-ordinates find something like their first modern use in the work of Newton and do not feature in *La Geometrié*. Boyer, *History of Analytic Geometry*, 26, 46.

[233] UKMT Senior Mentoring Scheme, 2014-2015, Sheet 4, Q5.

context of justification, with the latter now seeming more definitive.[234] Consider now a second example of analytic geometry in action, this time from complex analysis. We will see how algebraic methods enable us to be more explicit than when using visual intuition. We first define a convex subset of $\mathbb{C}$ as follows.

Definition 5.3. A subset $S \subset \mathbb{C}$ is convex iff $\forall a, b \in S$ the line segment connecting a and b is contained entirely within S.

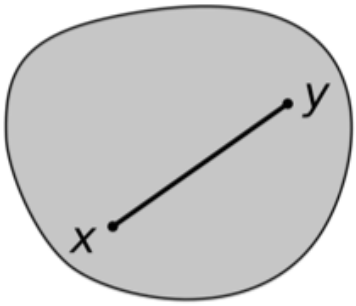

Figure 5.4. Diagram of a convex set containing elements x, y.
https://en.wikipedia.org/wiki/Convex_set

Now, consider the right-hand complex half-plane: the set $R = \{z \in \mathbb{C} \,|\Re(z) > 0\}$. The belief that this set is convex is easily supplied by intuition and visual imagination. Yet our intuitions are somewhat vague and indeterminate because there are so many kinds of cases. Are we sure we have included all of them in our reasoning? When this result is proved in elementary complex analysis courses, a more explicit algebraic-symbolic approach is preferred. We first identify the line segment from a to b as the following set of complex numbers:

$$\{at + b(1-t) \mid t \in [0,1]\}$$

Now, let $a, b \in R$ and let z lie on the line segment between a and b. We aim to show $z \in R$.

If $z = a$ or $z = b$ we are done. Suppose not. Then $z = at + b(1-t)$ for some $t \in (0,1)$. We now have that

$$\Re(z) = \Re\big(at + b(1-t)\big) = t\Re(a) + (1-t)\Re(b) > 0$$

so that $z \in R$ as required.

[234] Israel Kleiner, "Rigor and Proof in Mathematics: A Historical Perspective", *Mathematics Magazine* 64 (1991): 301.

This proof does not rely on intuitions about the geometry of a flat planar surface. Moreover, the formulation of a line as a set of points satisfying an algebraic condition that is expressed algebraically also makes the corresponding definition of convexity more exact as well. Another intuitive result for which explicit proof is required is the Jordan Curve Theorem: that the removal of a non-intersecting continuous closed curve from the plane splits it into two connected components, one of which is bounded. Though the result again seems visually obvious, this time an explicit proof is rather more difficult to find, and some special cases such as fractal curves like Koch's Snowflake require considerable effort to deal with. The non-obviousness of the result was pointed out by Bolzano, who also gave the first precise formulation of the problem.[235]

As well as these developments in analytic geometry, there is a parallel history of the decline of intuition in synthetic geometry too. We mention just two key developments. Following millennia of failed attempts to prove Euclid's parallel postulate from his other axioms, mathematicians such as Bolyai and Lobachevski came to explore the consequences of dropping the postulate, ultimately resulting in the study of a variety of non-Euclidean geometries. The differences between these were later described in algebraic (specifically, group-theoretic) terms. Intuitions based on everyday physical thinking about space, which largely seem to conceive of it as Euclidean, are of less help when working with these alternative geometries, and a more explicit, axiomatic approach is preferable.

A further departure from methods based on spatial intuition was made with Hilbert's 1899 *Der Grundlagen der Geometrie*. In this influential work, Hilbert built up synthetic geometry axiomatically from first principles. Though his text does contain diagrams, the proofs are fully explicit and do not rely essentially on visual intuition: 'we will use figures often. However, we will *never rely on them*.'[236] This approach was influenced by an 1882 work by Moritz Pasch, who wrote that 'the theorem is truly demonstrated only if the proof is completely independent of the figure.'[237]

Hilbert's axiomatic approach is so Abstract and Explicit that it is unnecessary even to give interpretations of the entities his system concerns. He begins with three kinds of objects – points, lines, and places – but as he later remarked, these could equally well be replaced by 'tables, chairs and beer mugs'.[238] Indeed, the concepts of 'line' and 'point' can often be simply interchanged, giving corresponding 'dual' theorems. We have thus moved a long way from Euclid's familiar and intuitive semantic definitions, such as of a point as 'that which has no parts'.

Speaking more generally, over the past few centuries mathematics has moved away from complex intuitive leaps in the context of justification, and towards explicit derivations where every step is clearly shown. In the next section, we will

---

[235] Fiona Ross and William Ross, "The Jordan Curve Theorem is Non-Trivial", *Journal of Mathematics and the Arts* 0 (2009): 3.
[236] Quoted in Paolo Mancosu, "Visualization in Logic and Mathematics", in *Visualization, Explanation and Reasoning Styles,* ed. Paolo Mancosu, Klaus Frovin Jørgensen, and Stig Andur Pedersen (Dordrecht: Springer, 2005), 14
[237] Quoted in ibid., 14.
[238] Constance Reid, *Hilbert* (New York: Springer, 1996), 60.

focus on the conceptual clarity of modern analysis and how its central objects are attended by clear necessary and sufficient defining conditions. Here we will also see how both the concepts and the modes of justification their use permit also became logically divorced from the spatiotemporal intuitions by which they may have originally been arrived at. Our intuitive concepts such as 'curve' and 'tangent' will turn out to be too vague for progress to continue, so that now 'intuition cannot be considered final in analysis'.[239]

Visual techniques have however made a comeback in recent years and are now highly important for discovering results in differential geometry, topology and complex analysis. This is in part due to the development of the computer: practitioners are able to plot graphs of entities like the Mandelbrot set with a level of detail that would have previously been unthinkable (see also Section 2.iii). However, these approaches must always be supplemented with rigorous logical proof and are never final in the context of justification.[240] This Intellectual Virtue of contemporary mathematics we have called Explicitness.

We should again pause here to make a brief caveat regarding our general thesis about contemporary mathematics. It is not that such explicit proofs are undergone each time a result such as the triangle inequality for complex numbers – also easily supplied by visual intuition – is given; but rather, that a result which lacked any publicly available, explicit proof would now be regarded as problematic. As in the previous section, we are talking about a modal constraint here.

Lastly, we close the section by briefly noting some further advantages of Explicitness within contemporary mathematics. We saw a brief glimpse of one such advantage for geometry in the discussion of Problem 5.2. Here our algebraic approach to the problem had a certain mechanicalness whereby very little creative imagination was required to find the answer. The application of algebraic methods can bring problems that may once have required the genius of an Archimedes within the reach of an average student; a deep intuitive grasp of the problem is now often unnecessary because a basic facility with symbolic manipulation suffices to reach the solution.

Two other advantages of Explicitness have already been explored in Chapter 2, where we saw proof was in many cases superior to non-deductive methods, both epistemically and in securing agreement amongst practitioners. If we are to rely on impenetrable leaps of intuition in the context of justification, we have no *a priori* guarantee that different practitioners will agree in their judgements – this may then damage Consensus. Moreover, judgement based on visual intuition has repeatedly proved itself to be insufficiently reliable in many areas of mathematics, and especially in analysis (see for example the discussion of Weierstrass's nowhere-differentiable function in the next section).

[239] James Pierpont, "On the Arithmetization of Mathematics", *Bulletin of the American Mathematical Society* 5 (1899), 394.
[240] Mancosu, "Visualization in Logic and Mathematics", 18-20.

Lastly, in proving that the right-hand complex half-plane $R = \{z \in \mathbb{C} \,|\Re(z) > 0\}$ was convex we also saw that the demand for explicit derivations can drive progress towards clear definitions. This brings us directly to the next section.

5.v. Conceptual Clarity in Contemporary Mathematics

In this section, I claim that the concepts of contemporary mathematics always have clear, shared, publicly available, necessary and sufficient defining conditions. Likewise, specific objects discussed (such as the exponential function) are always given clear definite descriptions. The chief historical illustration of this thesis will be the rigorous formulation of real analysis that took place in the 19th Century, wherein the collection of powerful results and calculation techniques developed by Newton, Leibniz, Lagrange, Euler and others were given rigorous and systematic treatment by Cauchy, Bolzano, Weierstrass, Dirichlet and Abel.[241]

Newton discovered many powerful results in the general study of curves. By the mid-1660's, he was able to differentiate and integrate any polynomial expression, and to apply the same procedures to more or less all the algebraic and mechanical equations then known. He realised that differentiation and integration were inverse operations, what we now call the Fundamental Theorem of the Calculus, and went on to develop a wide range of integration techniques.[242] He also inquired into the foundations of his achievements, though this time with only limited success. He was able to avoid the paradoxical infinitesimals – 'neither finite quantities, nor quantities infinitely small, nor yet nothing'[243] – only by introducing other obscure concepts, such as prime and ultimate ratios, and nascent (just coming into being) and evanescent (about to cease to exist) quantities.

In the meantime, Leibniz was responsible for similar developments on the continent. His notation facilitated a more general conception of differentiation with respect to variables other than time, and the algorithmic nature of his techniques enabled them to be more readily applicable. Yet the foundational problems still remained. Lagrange later focused attention upon them by proposing the rigorous foundations of the calculus as a prize problem to the Berlin Academy in 1784. His lectures at the *École Polytechnique* were published in two books in 1797 and 1799-1801, and these efforts influenced both Bolzano and Cauchy.[244]

Lagrange hoped to base the calculus on the concept of a derivative. Yet rather than following Newton's geometrical approach he simply defined the derivative as the coefficient of h in the Taylor series expansion of the function $f(x + h)$. He thus demonstrated the same faith in symbolic formalism shown by Leibniz and Euler (who was his doctoral supervisor), and sought to ground analysis in what he thought was the secure basis of algebra.

---

[241] Boyer, *The History of the Calculus and its Conceptual Development*; Judith Grabiner, "Is Mathematical Truth Time-Dependent?"
[242] Isaac Newton, "The October 1666 Tract on Fluxions", in *The Mathematical Papers of Isaac Newton*, ed. D. T. Whiteside (Cambridge: Cambridge University Press, 1967).
[243] George Berkeley, *The Analyst*, in *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, ed. William Ewald, (Oxford: Oxford University Press), 81.
[244] Israel Kleiner, "Rigor and Proof in Mathematics: A Historical Perspective", 296.

However, like Newton before him, Cauchy ultimately believed that the algebraic approach was problematic. Instead he saw that the route to a successful systematisation of real analysis was to base it on the fundamental concept of a limit.[245] He would employ this highly successfully, and though giving only a verbal definition it was clear from his proofs – in which the definitions had featured prominently, as those of earlier mathematicians including Euler had not – that he had essentially our modern conception. In the 1870's, Weierstrass – who worked out the approach more systematically – replaced Cauchy's verbal definition with a symbolic definition using quantifiers. This is essentially the definition we use today:

Definition 5.5.

A sequence $\langle x_n \rangle$ tends to a real limit l as n tends to infinity iff:

$$\forall \varepsilon > 0,\ \exists\, N \in \mathbb{N} :\ n \geq N \longrightarrow |x_n - l| < \varepsilon.$$

Definition 5.6.

A real function f tends to a limit l as its argument tends to ∞ iff:

$$\forall \varepsilon > 0\ \exists\, x :\ \forall y, \qquad y > x \rightarrow |f(y) - l| < \varepsilon$$

in which case, we write $\lim_{x \to \infty} f(x) = l$.

The discovery of these definitions constituted a substantial mathematical achievement: they encompass everything bound up with our intuitive conceptions of limits using only functions, real numbers and logical quantifiers, with no vague references to space, time, motion, or infinite processes. Similar definitions are of course available in other cases, where the argument and value of the function tend either to specific real numbers or to $\pm\infty$. The definitions of the other central analytic properties of functions can now be given:

Definition 5.7.

A real function f is continuous at a point a of its domain iff:

$$\lim_{x \to a} f(x) = f(a)$$

Definition 5.8. The derivative of a real function at x is equal to the following limit, if it exists:

[245] Judith Grabiner, "Who gave you the epsilon? Cauchy and the origins of rigorous calculus", *The American Mathematical Monthly*, 90 (1983): 185– 194.

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

In this case, we say that f is differentiable at x.

Clearly, this latter definition takes its inspiration from Newton's geometric approach. For the kinds of well-behaved continuous functions mathematicians had largely dealt with before the 19th Century, the value of the expression considered is equal to the gradient of the secant, which on a graph of such functions visually appears to approach that of the tangent. Yet the definition itself is logically independent of geometrical considerations and is applicable to any function whatever, even those for which the idea of graphical representation is problematic. For instance, we define a function f as follows:

$$f(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right) & , \quad x \neq 0 \\ 0 & , \quad x = 0 \end{cases}$$

Then according to Definition 5.8 the function f is differentiable everywhere, despite its chaotic behaviour near the origin. We can also rigorously connect the property of differentiability with other attributes: if a function is differentiable at a point then it is always continuous there too, for example.

It is also clear from Definitions 5.5 – 5.8 that by the 19th Century focus had shifted from the general study of algebrao-geometric properties of curves to a study of the formal analytical properties of functions. We therefore also give a brief discussion of the function concept.[246] When functions first emerged as an explicit subject of study in the 17th Century, it was in connection with the development of analytic geometry and the calculus, and the physical applications being made of this new mathematics. For Bernoulli, a function was a mathematical expression composed of a variable and some constants that could be evaluated for any real input. Euler gave a similar definition in 1748:

> 'A function of a variable quantity is an analytic expression composed in any way whatsoever of the variable quantity and numbers or constant quantities.'[247]

Yet later Euler generalised this definition to one that 'encompasses all the ways in which one quantity can be determined in terms of others', and his definition was repeated by Lagrange, Lacroix and Cauchy.[248] As increasingly exotic functions gained acceptance with some writers, it eventually became unclear which class of entities were being discussed. Since around the start of the 20th Century the

[246] For a detailed discussion, see A. P. Youschkevich, "The Concept of Function up to the Middle of the 19th Century", *Archive for History of Exact Sciences* 16 (1976): 37-85.
[247] Quoted in Victor J. Katz, "Euler's Analysis Textbooks", in *Leonard Euler: Life, Work and Legacy*, ed. Robert Bradley and C. Edward Sandifer (Oxford: Elsevier, 2007), 214
[248] Lützen, "Between Rigor and Applications", 472.

favoured solution has been to first define functions in a very broad way – the 'ordered pairs' formulation – and then later restrict attention to functions having certain properties, such as being continuous, uniformly continuous, differentiable, Lipchitz, and so on. We can thereby attain a high degree of rigour whilst still admitting as true those theorems that only apply to specific kinds of functions.

This 'ordered pairs' definition of a function f from X to Y is given using set theory. We begin with any two sets X and Y, called the domain and codomain of the function respectively. Intuitively speaking, the domain X is regarded as the set of entities (they need not now be real numbers) that can be the inputs or values of the argument of the function, and the codomain Y is regarded as the set of possible outputs to which these inputs are assigned; this is indicated with the notation $f: X \to Y$. Yet in mathematical terms we need not refer to this intuitive interpretation in terms of a rule of assignment at all, however complex and heterogeneous; instead the definition of a function can be given as a subset of this Cartesian product $X \times Y$ as follows.[249]

Definition 5.9.

A function $f: X \to Y$ is a subset of $X \times Y$ such that $\forall x \in X, \exists y \in Y : (x, y) \in f$ and moreover such that if $(x, y_1) \in f$ and $(x, y_2) \in f$ then $y_1 = y_2$.

Again, the definition is completely explicit and does not rest on a merely intuitive acquaintance with the concept at any point. We can also now define a sequence with members in X as a function $f: \mathbb{N} \to X$.

A further area where clarification was required was with the notion of a real number itself, which was at the time still being understood in a vague geometric idiom using the number line. However, with only this geometric notion mathematicians were unable to rigorously prove three important theorems from Cauchy's *Course d'analyse*: that a continuous function on a closed bounded interval is Riemann integrable, the Intermediate Value Theorem, and that an increasing sequence of real numbers that is bounded above converges to a real limit. This latter property is equivalent to the least upper bound or 'completeness' property of the real numbers: every non-empty subset of $\mathbb{R}$ that is bounded above has a least upper bound. It is also equivalent to the statement that every Cauchy sequence converges to a real limit. Completeness is of central importance in real analysis, and indeed its appearance is often taken to be a condition that differentiates the subject from what is otherwise mere algebra.[250]

Cauchy's arguments for the three theorems just mentioned rested on intuitive geometric reasoning. To further divorce the new mathematical analysis from spatial intuitions – which had by then proved unreliable – many mathematicians recognised that an explicit account of the real numbers in non-geometric terms

[249] We define the Cartesian product $X \times Y$ to be the set of all ordered pairs from X and Y, i.e. $\{(x, y) | x \in X, y \in Y\}$.

[250] Though it need not appear in a more general enquiry into analysis that is not specifically concerned with the real numbers. See later in this section.

was needed. This may be achieved by a combination of synthetic and constructive methods: we first give a system of axioms that specify the desired properties of real numbers as a totally ordered field together with the completeness property, and then construct an explicit model that can be shown to satisfy these axioms.

Various such models are now available; for instance, the decimal expansion of a real number enables us to associate it with an infinite sequence of decimal digits.[251] A second and historically prior construction is due to Cantor, based on ideas from Weierstrass. Cantor associated real numbers with sets of 'fundamental sequences' – what we would now call rational Cauchy sequences – which converge to the same real limit.[252] A third approach due to Dedekind[253] makes use of the concept of Dedekind cuts, which are partitions of the rational numbers into two sets A and B such that the following conditions are met:

i.) $\forall a \in A, b \in B : a < b$
ii.) $\forall a \in A, \exists a' \in A : a' > a$

All three of these constructions permit us to carry out arithmetic operations, and we can easily check that they satisfy the usual axioms of the real numbers – crucially, including the LUB property. Weierstrass' project of trenchantly purging all of analysis of the use of infinitesimals and intuitive geometrical reasoning can then be satisfactorily completed.

Another advantage of Weierstrass' approach is that the formal definition with logical quantifiers enables a greater appreciation of the importance of their order. This led to the rectification of some earlier omissions that Cauchy's merely verbal definition led him to commit. In particular, Cauchy did not distinguish between pointwise and uniform continuity, nor pointwise and uniform convergence of a sequence of functions:

Definition 5.10.

A real function f is continuous iff

$$\forall x \, \forall \varepsilon \in \mathbb{R}^+ \; \exists \delta \in \mathbb{R}^+ : \; \forall y, \;\; |x - y| < \delta \;\; \rightarrow |f(x) - f(y)| < \varepsilon$$

And uniformly continuous iff

$$\forall \varepsilon \in \mathbb{R}^+ \; \exists \delta \in \mathbb{R}^+ : \;\; \forall x \, \forall y, \;\; |x - y| < \delta \;\; \rightarrow |f(x) - f(y)| < \varepsilon$$

[251] Care must be taken to avoid dual representation – for example, $1 = 0.99999\ldots$. See Timothy Gowers, "What is so Wrong With Thinking of Real Numbers as Infinite Decimals?", *Department of Pure Mathematics and Mathematical Statistics*, accessed 19th August 2015. https://www.dpmms.cam.ac.uk/~wtg10/decimals.html

[252] Marcus Giaquinto, *The Search for Certainty: A Philosophical Account of Foundations of Mathematics* (Oxford: Oxford University Press, 2002), 17.

[253] Ibid., 17-18.

Definition 5.11.

A sequence of real functions $\langle f_n \rangle$ converges pointwise to a function f iff:

$$\forall x \forall \varepsilon \in \mathbb{R}^+ \, \exists N \in \mathbb{N} : \ \forall n \in \mathbb{N}, n \geq N \rightarrow \ |f_n(x) - f(x)| < \varepsilon$$

And converges uniformly to a real function f iff:

$$\forall \varepsilon \in \mathbb{R}^+ \, \exists N \in \mathbb{N} : \ \forall x \, \forall n \in \mathbb{N}, n \geq N \rightarrow \ |f_n(x) - f(x)| < \varepsilon$$

Conflating both pairs of concepts led Cauchy to arrive at the false result that a convergent series of continuous functions is a continuous function – a claim he took himself to have proved, using infinitesimal methods.[254] Abel pointed out a counterexample in 1826. The following series of continuous functions:

$$f_n(x) = \frac{\sin(x)}{1} - \frac{\sin(2x)}{2} + \frac{\sin(3x)}{3} - \ \ldots + (-1)^{n+1} \frac{\sin(nx)}{n}$$

converges to a function that is discontinuous at $x = (2n + 1)\pi$ for all integers n.[255] The correct result – the Uniform Limit Theorem – requires the convergence to be uniform, whereas for this series it is only pointwise. For some other purposes, stronger conditions such as 'equicontinuity' are also needed.

We now discuss some further advantages of this new approach based on clear definitions rather than a merely intuitive acquaintance with the underlying concepts. One consequence is that definitions come to feature more prominently in the proofs, as opposed to merely vaguely indicating a subject matter for discussion. Proofs then become far easier to construct and again are almost mechanical: hence it is often remarked that analysis proofs seem to 'write themselves'. More generally, a solid grasp of the underlying definitions makes writing proofs of elementary results easy: often we need only write out what is to be proved as explicitly as possible and see the connections emerge.

Another consequence of Univocality is that the logical structure of proofs becomes more transparent, as we can see exactly which properties of functions or sequences are being used. We have already noticed the converse of this claim in the previous section; indeed it seems plausible to think that the quality of arguments and the adequacy of definitions will tend to be correlated.[256]

Next, once it is clear what is essentially involved in the concept of a real limit, it then becomes easier to generalise to other cases. From the symbolic definition we see that all that is really needed is a set equipped with a notion of distance. This

[254] Kleiner, *Rigor and Proof in Mathematics*, 299.
[255] Umberto Bottazini, *The Higher Calculus* (New York: Springer-Verlag, 1986), 113.
[256] Again see Lakatos, *Proofs and Refutations*.

soon led to the extension of the limit definitions to the complex numbers, with the complex modulus function replacing the absolute real value. The other definitions given above extend similarly, and the notion of an integral can also be analogously applied; this has led to the fascinating theory of complex analysis. Moreover, not only the theorems but also the proofs are often identical (though other results are highly surprising and counterintuitive).

Eventually, mathematicians further generalised complex analysis to consider functions defined on arbitrary metric spaces X and Y. Not only do the above definitions generalise accordingly, but also many of the theorems from real analysis also have analogues within this more general approach, and the proofs are again often very similar. Consider the following result from real analysis:

Theorem 5.12.

Let f be a continuous function on a closed bounded interval $[a, b]$. Then f is bounded and attains its supremum and infimum.

We can also prove the following generalisation:

Theorem 5.13.

Let f be a continuous function on a compact metric space X. Then f is bounded and attains its supremum and infimum.

Since the 19th Century, this requirement of clear defining conditions has become characteristic of mathematical research more broadly. There is a parallel history of the development of abstract algebra over roughly the same period, initially motivated by developments in the theory of equations, and the strikingly original work of Évariste Galois. George Peacock introduced axiomatic thinking into arithmetic and algebra, and De Morgan, Gibbs and Cayley also produced influential work that switched attention to the study of highly general abstract algebraic systems such as groups, fields, and vector spaces.[257]

This approach based on abstract concepts attended with clear definitions gives a wonderfully economic method of treating many systems at once. For instance, if a theorem is proved about a general class of structures, such as vector spaces $(V, \mathbb{F}, *, +)$, it automatically applies to a wide variety of specific systems, such as n-dimensional vectors with elements from a field $\mathbb{F}$, the set of matrices M of any dimensions with elements from $\mathbb{F}$ under matrix addition and scalar multiplication, the space $V = Hom(U, W)$ of linear maps between vector spaces U and W under composition and scalar multiplication, the space of continuous functions on a closed bounded real interval $[a, b]$ with pointwise addition and scalar multiplication, and many more besides. Moreover, beyond pragmatic questions of economy, the interest often lies in the connections that are made clear between such structures when what is essential to them has been isolated.

[257] For an overview see Merzbach and Boyer, *A History of Mathematics*, chapter 21.

Within this new structuralist approach to mathematics, wherein only the most general features of different kinds of objects and systems are given, merely intuitive understanding becomes less effective. A general familiarity with the system, such as physical intuition provides for some continuous real functions of one variable, is not available because we will usually not even know specifically which system we are talking about. And many of the underlying concepts are far removed from everyday experience, so that our intuitions about them cannot be relied upon in the absence of clear definitions. During a lecture delivered by von Neumann in Göttingen in the 1920's, about linear operators on Hilbert Spaces, Hilbert – who was in the audience – was alleged to have stood up and asked 'Yes, Herr von Neumann, but what actually is a Hilbert space?'[258]

The Univocal formulation of a concept can also aid Consensus. For instance, there is now no trace of the long controversy about what is to count as a function. Moreover, once we are clear about definitions, theorems of impressive strength and sophistication can be proved – but prior to this, such achievements are often not available to us. Consider trying to arrive at (or even state) Theorem 5.13 without having clear definitions of the concepts 'continuous', 'function', 'compact', 'metric space', 'bounded', 'supremum' and infimum'. Each concept plays an important part and is essential to the truth of the theorem: if any of the antecedent conditions are not met, then the consequent may not be true.

Once such clear definitions of the important concepts are available, mathematics becomes the study of patterns of idealised deductive relationships within structures.[259] The precision of theorems in modern abstract mathematics means they can be confidently applied without exception to all eligible systems. If this were not the case, the structuralist approach would be far less effective. We would always need a separate enquiry in each case to see whether a system we were interested in which apparently fell under the scope of a theorem really did meet the conditions of its application, and whether the conclusion really did follow. Within modern abstract mathematics such moves are unnecessary: we only need check that we do indeed have an instance of a group, ring, field, vector space, metric space, or whatever class of entities the theorem we wish to make use of concerns.

Again, before moving on we must qualify the claim that all mathematical concepts actually in use meet the standard of Univocality. For example, consider the follow definition from Stein and Sakarachi's *Complex Analysis*: 'We call a toy contour any closed curve where the notion of interior is obvious'.[260] This concept also appears in the statement of a restricted version of Cauchy's Theorem.[261] Yet just as a person possessing the virtue of honesty may sometimes lie on occasion, the normative ideal of necessary and sufficient defining conditions is present even in those cases where it is not attained. Mathematicians will typically see such counterexamples as areas where more progress is required.

---

[258] Kleiner, *Rigor and Proof in Mathematics*, 303.
[259] Again, see Corfield, *Towards a Philosophy of Real Mathematics*, 181.
[260] Elias Stein and Rami Shakarachi, *Complex Analysis* (Princeton: Princeton University Press, 2003), 40.
[261] Ibid., 41.

Lastly, there are some concepts such as the membership symbol $\in$ in set theory that are taken as fundamental and so do not require explicit definition. Though the meaning of these concepts is nevertheless clear, it is only indicated through the use of examples.

5.vi. Formalizing Mathematics

In this section, we see how during the early 20th Century mathematicians began formalizing their mathematical arguments. This required them to make explicit their mathematical and logical axioms, and the rules of inference used to derive results. These techniques also allow us to show with the minimum of assumptions that the rules of inference mathematicians make use of are truth-preserving (c.f. Section 1.1). Moreover, any argument that is not translatable into such a system would now be considered problematic.

The story of this development continues from the previous section. Once a construction of the real numbers – such as Cantor's or Dedekind's – has been given, we are still faced with two problems. The first is to supply a corresponding account of the rational numbers, on which such definitions rest. The second is that the constructions of the real numbers given above make use of infinite sets whose members cannot be given explicitly. Cantor's account refers to the set of all fundamental sequences and hence the set of all infinite rational sequences of which it is a subset, and Dedekind's account assumes a set of all cuts.

Rational numbers can be interpreted as equivalence classes of ordered pairs of integers,[262] and so the first problem further reduces to giving an account of the integers and then the natural numbers. One such account was given by Frege, who identified the natural number n with the class of all classes containing n elements. He also interpreted classes as the extensions of predicates.[263] Russell later showed that this approach to set theory is problematic by considering a predicate Fx defined to mean that x is not a member of itself. But there can be no class that forms the extension of this predicate: otherwise it would have to both contain itself and not contain itself. This is Russell's paradox, which shows that Frege's system was inconsistent, as was revealed to Frege the night before publication of the second volume of his efforts in 1902.[264] Two further set-theoretic paradoxes had also been discovered previously: the Burali-Forti paradox and Cantor's paradox.[265]

Whilst Russell and Whitehead were attempting to fix the issues with their understanding of set theory in terms of classes conceived as extensions of predicates, Zermelo developed an alternative axiomatic approach. His 1908 paper, he gave a set of axioms for set theory that enabled the development of arithmetic and analysis whilst avoiding the known paradoxes.[266] But although this was an

[262] The classes form a partition of the set $\{(a,b)|a \in \mathbb{Z}, b \in \mathbb{N}\}$ under the equivalence relation R, where $R((a,b),(c,d))$ iff $ad = bc$.
[263] Frege, *The Foundations of Arithmetic*. See also Giaquinto, *The Search for Certainty*, 33.
[264] Giaquinto, *The Search For Certainty*, 53.
[265] Ibid., Part II.
[266] Ernst Zermelo, "Untersuchungen über die Grudnlagen de Mengenlehre", *Mathematische Annalen* I (1908): 261-281.

important achievement, doubts about his system were not eliminated entirely. The presence of further sources of inconsistency had not been definitively ruled out, and still seemed an open possibility. The same was true of Russell and Whitehead's modified system.

With these issues in mind, Hilbert aimed to prove using only finitary methods[267] that classical mathematics was reliable in the sense that it could not be used to derive false finitary consequences. As these finitary methods were highly transparent, and so possessed a particularly high degree of epistemic security, the completion of this project would have served to remove any remaining doubts about classical mathematics. His approach to achieving this goal involved formalising set theory in a way that was precise enough for the formalized system, together with some logical axioms, to itself become the subject of mathematical investigation.

Formal systems in this sense consist of a completely specified formal language together with an effectively decidable set of axioms in that language and a finite set of inference rules. Sentences of the language are composed of finite strings of symbols from a fixed alphabet. The specification for allowable sentences gives explicit rules of composition such that we have an effective way of deciding, for any string of symbols, whether it is a sentence in the language. Sentences are arranged into sequences called derivations, such that each line in the derivation is either an axiom (either logical or mathematical) or follows from one or more earlier lines by inference rules (e.g. Modus Ponens) that are also given explicitly. For any given sequence of sentences, we can thus always in principle effectively determine whether it is a derivation.

We now state the soundness theorem for first-order predicate logic with identity. Within a formal language L we can define both semantic (i.e. truth-theoretic) and syntactic (i.e. proof-theoretic) notions of entailment. For a set of sentences Γ in L and a single sentence θ in L, the first notion, written $\Gamma \vDash \theta$, says that θ is true for every interpretation $\Im$ of L in which each member of Γ is true. The second notion, written $\Gamma \vdash \theta$, says there is a derivation of θ using only the elements of Γ as axioms (in addition to the logical axioms). The soundness theorem for first-order predicate logic then says that $\Gamma \vdash \theta \Rightarrow \Gamma \vDash \theta$. Moreover, the proof relies on very little mathematics, all of which is highly uncontroversial.[268]

The soundness theorem for first-order logic entitles us to be confident that any mathematical argument that can be formalised as a purely deductive proof from explicit premises is valid. The possibility of recasting mathematical arguments in accord with such a model thus became a widely accepted standard, and any argument that cannot be formalised in this way is now counted as flawed or at least unclear. This is the Intellectual Virtue we have called 'Formalizability'.

We now discuss some of the advantages of Formalizability for contemporary research. One such benefit has already been made clear: this practice provides an

[267] For the definition of 'finitary' subject matter, see Giaquinto, *The Search for Certainty*, 145-146.
[268] The converse result, Gödel's Completeness Theorem, was proved in 1929. Kurt Gödel, "Über die Vollständigkeit des Logikkalküls", *Doctoral Dissertation* (University of Vienna, 1929).

effective check on the validity of our arguments and of accepted mathematical rules of inference (again, see Section 1.1). Secondly, it also gives a check on Univocality: in attempting to formalise our arguments we may realise that some of our definitions are not sufficiently clear.

Formalizability also enables automated proof checking and the creation of large online libraries of proofs. One such attempt was the QED project, an attempt to build a 'single, distributed, computerized repository that rigorously represents all important, established mathematical knowledge.'[269] Its founders hoped to improve the Reliability of Publicly Accepted mathematics, and to facilitate smoother collaboration between research mathematicians, who would be able to quickly and efficiently scan the database for results relevant to their research.[270]

Sadly, the QED project did not really get off the ground and was abandoned in 1996. The same idea has since been implemented successfully, however: in the Mizar system for instance. This also consists of a formal language (based on Tarski-Grothendieck set theory), a program for mechanically checking proofs written in this language, and a library of formalized mathematics encompassing some 49,000 theorems.[271] Other such projects also exist, including the Metamath program,[272] and verification programs are now numerous.[273] The proofs of many important results have been successfully formalized, including the Prime Number Theorem,[274] the Hahn-Banach Theorem,[275] Brouwer's Fixed Point Theorem and the Jordan Curve Theorem,[276] the Robbins Conjecture,[277] Gödel's Incompleteness Theorems,[278] and (as mentioned in Section 1.vi) the Four Colour Theorem.[279]

Formalizability can thus provide an effective alternative solution to the problems discussed in Section 1.vi: that proof presentations are sometimes so long that they tend to thwart the checking process, and that some proofs essentially make use of computers. Although computers are still used to automatically check that each step in the formalised proof is valid, these checks are simple and highly transparent – unlike the original 1976 proof of the Four Colour Theorem, where the computer was programmed to carry out complex chains of reasoning.

---

[269] "Summary", QED. Accessed 19th August 2015. http://mizar.org/qed/

[270] "The QED Manifesto", in "Automated Deduction – CADE 12", *Springer-Verlag*, *Lecture Notes in Artificial Intelligence* 814 (1994): 238-251.

[271] "Mizar Project", accessed 18th August 2011. http://mizar.org/project/

[272] http://metamath.org

[273] For instance, ACL2, Coq, HOL Light, HOL4, ProofPower, IMPS, Isabelle, Mizar, NUPRL and PVS.

[274] Jeremy Avigad, Kevin Donnelly, David Gray, Paul Raff, "A Formally Verified Proof of the Prime Number Theorem", *ACM Transactions on Computational Logic* 9 (2007): 1-23.

[275] Gertrud Bauer and Markus Wenzel, "Computer-Assisted Mathematics at Work (The Hahn-Banach Theorem in Isabelle/Isar)", in *Selected Papers from the International Workshop on Types for Proofs and Programs*, ed. Thierry Coquand, Peter Dubjer, Bengy Nordström and Jan M. Smith (London: Springer-Verlag, 2000), 61-76.

[276] Artur Korniłowicz, "A Proof of the Jordan Curve Theorem via the Brouwer Fixed Point Theorem", *Mechanized Mathematics and its Application* 6 (2007): 33-40.

[277] William McCune, "Solution of the Robbins Problem", *Journal for Automated Reasoning* 19 (1997): 263-276.

[278] Larry Paulson, "A Machine-Assisted Proof of Gödel's Incompleteness Theorems for the Theory of Hereditary Finite Sets", *Review of Symbolic Logic* 7 (2014), 484-498.

[279] Georges Gonthier, "Formal Proof – The Four-Color Theorem", *Notices of the AMS* 55 (2008): 1382-1393.

Moreover, it is also possible to translate formalised proofs into a special self-correcting 'probabilistically checkable' format. Probabilistic checking of formal proofs enables us to 'become confident of their validity by checking only 10 or 20 bits chosen randomly in a correlated way.'[280] The Probabilistically Checkable Proofs Theorem, a result of immense scope and value, tells us that '*any* mathematical theorem, in any standard formal system such as Zermelo-Fraenkel set theory, can be converted in polynomial time into a probabilistically-checkable format.'[281]

The use of these techniques would not require us to revise the traditional rule of proof prior to publication, as they remain a purely external check on the validity of a deductive argument and do not form an essential part of it – just as Euler supplemented his theoretical work on infinite series with numerical checks. Like the hand-held calculator, these techniques could improve the Reliability of mathematics without interfering with the underlying logic of mathematics itself.

Again, a brief caveat before moving on. At the moment, it is not true that every Publicly Accepted mathematical truth has been successfully formalized. However, any argument that essentially relies on a rule of inference that can be given no interpretation in such a system – and hence cannot be checked to be truth-preserving – would surely now be treated with suspicion. Like Abstractness, Explicitness and Univocality, Formalizability is thus also a modal condition.

5.vii. Conclusion

In this chapter, we have articulated four standards of excellence that apply to published mathematical discourses, and given brief justifications for retaining them in the future. These were Abstractness, Explicitness, Univocality and Formalizability. In the next chapter, I will argue that Probabilistic Arguments cannot adhere to these four standards, thus providing reasons for mathematicians to reject them in the context of Public Acceptance.

We have also mentioned a highly effective and widely applicable technique for improving the Reliability of Publicly Accepted mathematics, based on translating formal proofs into a special error-correcting format. This may constitute an alternative remedy to the problems with proof outlined in Section 1.vi, and one that will not require mathematicians to Publicly Accept mathematical arguments that do not comply with the four Intellectual Virtues.

---

[280] Scott Aaronson, "Why Philosophers Should Care About Computational Complexity", in *Computability: Turing, Gödel, Church, and Beyond,* ed. B. Jack Copeland, Carl J. Posey, Oron Shagrir (United States: MIT Press, 2013), 302

[281] Ibid., 302.

6. Mathematics and Probability

In the opening section of this chapter I give a clear example of how the Intellectual Virtues (Abstractness, Explicitness, Univocality and Formalizability) can supply us with reasons for rejecting arguments for mathematical claims in the context of Public Acceptance. I then ask whether the arguments of Chapter 4, which established the epistemic superiority of the Rabin-Miller Algorithm to the Trial Division Algorithm in the context of Private Acceptance, carry as much weight when we consider Public Acceptance. In the rest of the chapter, I argue that to allow Probabilistic Arguments in the context of Public Acceptance would lead to the Intellectual Virtues being undermined.

6.i. Public Acceptance and Non-Mathematical Arguments

We open with an example of how the normative principles articulated in the previous chapter can give us reasons to exclude certain modes of enquiry from being part of Publicly Accepted mathematics, even if the conclusions arrived at through them are of a distinctly mathematical kind. A means of justifying a mathematical statement may still constitute an unacceptable attempt at mathematics even if it employs a method that can be shown to be highly reliable.

The reader may recall that in Section 2.i we briefly mentioned the DNA computer methods of Leonard Adleman, which can be used to solve problems in pure mathematics.[282] One such problem is that of finding a Hamiltonian path in a given directed graph on n vertices.[283] Finding a Hamiltonian path directly through combinatorial analysis can become computationally intensive for large values of n, but the DNA algorithm is usually able to find one much more efficiently. Moreover, if the algorithm does not find one then there is a very strong chance that no such path exists.

DNA molecules have a double-helix structure with two coils of proteins, each either guanine, adenine, thymine and cytosine, denoted herein by the letters G, A, T and C. G will only bond with C and conversely, whereas A will only bond with T and conversely. So given a string of proteins – such as A-T-C-G-A – there is a unique string of that length which will bond with it at every place – i.e. T-A-G-C-T. The algorithm exploits this property as follows. Firstly, we represent each town by a unique 20-molecule long strand of DNA. We then also represent all the one-way roads (x, y) as 20-molecule long strands DNA by taking the last 10 molecules from the formula for the town x, and the first 10 from the formula for the town y. When the strands of DNA are allowed to bond together, each road strand of DNA

[282] Leonard Adleman, "Molecular Computation of Solutions to Combinatorial Problems", *Science* 266 (1994): 1021-1024. See also Don Fallis, "The Epistemic Status of Probabilistic Proof", *The Journal of Philosophy* 94 (1997): 165-186.

[283] A directed graph is such that the edges are ordered rather than unordered pairs of vertices: imagine n towns joined by one-way roads. A Hamiltonian path on a graph of size n is a collection of adjacent edges $\{(x_1, x_2), (x_2, x_3), \ldots, (x_{n-1}, x_1)\}$ such that each x_i is distinct: that is, a route that visits each town exactly once and returns to the starting point.

act as a kind of 'molecular splint' that joins a pair of towns that are adjacent on the original graph. The diagram illustrates the idea with molecules of length 6 instead of 20:

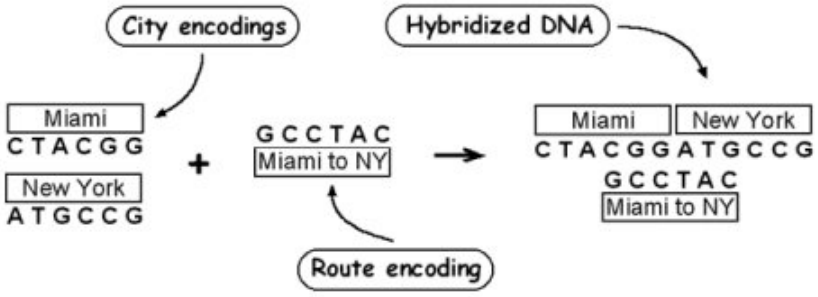

Diagram 6.1. A depiction of the binding of two vertices using DNA.

After paths of various lengths have been allowed to form, the next stage is to sort the strands by length using a gel separation process. Only strands of length $20n$ are retained, with the rest being discarded. Then we heat the mixture until the double helixes break apart into single strands of DNA. Next, magnetised copies of the inverse of the strand representing the first town in the graph are added to the tube. A magnet is held against the container, isolating all the strands that correspond to paths visiting the first town. The rest of the strands are discarded, and the mixture is heated to split the remaining strands from their magnetised inverses. This process is repeated for each town so that only strands corresponding to paths of length n that visit every town remain.

If any strands of DNA remain at the end of the process, we can easily check against the original graph that they do in fact correspond to Hamiltonian paths. If, on the other hand, no strands remain, then the chances of there being a Hamiltonian path in the graph is very slim. Adleman says in the instructions for implementation that 'the quantity used should be just sufficient to ensure that during the ligation step (step 1), a molecule encoding a Hamiltonian path will be formed with a high probability if such a path exists in the graph.'

As before, if this probability is high enough then Private Acceptance that there are no Hamiltonian paths may under some circumstances be rationally required. Yet it is intuitively clear that such a procedure is unsuitable for establishing this claim as a piece of Publicly Accepted mathematics, and that mathematicians would be warranted in rejecting it in this context. Though the conclusion is a mathematical proposition, the argument contains further propositions that essentially refer to physical properties of DNA, contravening Abstractness. These are only known empirically, and no clear derivation can be given, flouting Explicitness. The argument also involves physical concepts not amenable to precise definition in a way that would satisfy a mathematician, and so fails to meet Univocality. Lastly, the argument is certainly not Formalizable.

In the remainder of this section, I will use an analogy to show that the arguments from earlier chapters about the practical superiority of Probabilistic Arguments in the context of Private Acceptance have less purchase in the context of Public Acceptance. Here, not only the standards of a general individual epistemology but also the Intellectual Virtues are in play. We proceed by again comparing two algorithms for the same task: one of which will be more effective in practice, but nevertheless clearly inferior *qua* piece of mathematics.

In Chapter 2, we said that Gauss has proved that the class of compass and straightedge constructible polygons is exactly the class of n-gons for which n is the product of a power of 2 and any number of distinct Fermat primes.[284] For example, if $n = 17$ then the n-gon (called a heptadecagon) is constructible. The first such construction was given by Johannes Erchinger in 1825.[285] We give a different method, using Carlyle Circles.

Algorithm 6.2 (Constructs an heptadecagon of unit side length)

---

1. Given a line segment OA of unit length, construct a circle γ_1 centred at O passing through A. This circle will be the circumcircle of the heptadecagon. Produce OA to give a line meeting the circle γ_1 again at B. Construct the perpendicular to OA through O meeting γ_1 again at two points – call either one of these C.

2. Now bisect OB at the point D and construct a circle γ_2 centred at D passing through the point A. Construct also a line through D parallel to OC and let E be the point of intersection of this point with γ_2 that lies on the other side of OA as C.

3. Now draw a circle γ_3 centred at E passing through the point C. This will intersect with the line OA produced at two places. Let F be the intersection on the same side of OC produced as B, and let the other intersection of the circle γ_3 with OA produced be at the point G, so that G is on the same side of OC produced as A.

4. Now bisect OG to find the point H. Draw a circle γ_4 centred at H passing through C. This will meet OA produced at two points. Let I be the point on the same side of OC produced as A.

5. Now bisect OF to give the point J. Construct a circle γ_5 centred at J passing through C. This circle γ_5 will also meet the line OA produced at two points. Let K be the point of intersection of γ_5 that lies on the same side of OC as A, i.e. between O and A.

---

[284] To Recap, a Fermat prime is a prime of the form $2^{2^n} + 1$

[285] Karin Reich, "Die Entdeckung und frühe Rezeption der Konstruierbarkeit des regelmäβigen 17-Ecks und dessen geometrische Konstruktion durch Johannes Erchinger (1825)", in *Mathesis. Festschrift zum siebzigsten Geburtstag von Matthias Schramm,* ed. R. Thiele (Berlin: Diepholz, 2000), 101-118.

6. Now bisect AK to find the point L and draw a circle γ_6 centred at L and passing through O. Let the second intersection of this circle with the line segment OA produced be at M. Now draw a circle γ_7 centred at O and passing through M, meeting the line segment OC extended at N, where N is on the same side of OA extended as C (i.e. so that N is slightly further along OC from O than C is).

7. Now join N with the point I and bisect IN to find the point P. Construct a circle γ_8 centred at P passing through C. This will meet OA produced at two points: let Q be the point furthest from O (i.e. the one that is further along OA produced from O than A is). Now draw the perpendicular bisector of OQ, which will meet the original circle γ_1 at two points. Let one of these points be R – on the same side of OA as C, say. Then angle $A\hat{O}R$ is equal to $\frac{2\pi}{17}$ and AR is the side length of a heptadecagon centred at O with vertices at A and R. Mark 15 more points equally spaced around the circle and join these up to obtain the heptadecagon.

This algorithm is rather complex, and although other algorithms for the construction are available unfortunately none is significantly simpler than this one. We can however give a much simpler algorithm for producing a polygon with 17 sides that is very close to being regular. Consider a heptadecagon of unit side length. Using trigonometry, we have that the length of the radius of a circumcircle is equal to $\frac{1}{2}\operatorname{cosec}\left(\frac{2\pi}{34}\right) = 2.721095575876\ldots$. Now, we proceed as follows. Writing the fractional part in binary, we see it is equal to 0.1011100010 to ten binary decimal places. If we can construct a line segment of approximately this length, it will be correct to over one part in a thousand – well below the threshold of visual perception. Consider then the following algorithm.

Algorithm 6.3 (Constructs an approximate heptadecagon of unit side length)

Draw a line segment S of unit length, and then bisect the original length repeatedly to form the smaller lengths $x_1, x_2, \ldots, x_9$. Then join twice the original length S together with the sum $x_1 + x_3 + x_4 + x_5 + x_9$. Call this total length L. Construct an isosceles triangle with two sides equal to L upon the original line segment S, with its third vertex at the point O. Now construct a circle with a centre at O through the two endpoints of S. Use the unit length to mark 15 more points around the circle. Join up vertices that are adjacent on the circle to obtain a shape with 17 sides that are approximately equal.

I have used Apollonius to construct heptadecagons using both methods – though only going to the 5th binary decimal place rather than the 10th as suggested above, as the program was unable to accommodate this resolution (indeed, there seems something rather perverse about going that far, like rounding to too many decimal places after an approximation). The results are included below. I have actually forgotten which polygon is which, though the reader is challenged to guess (as we

effectively rounded the value down, one side will be shorter than the others on the polygon that is only approximate).

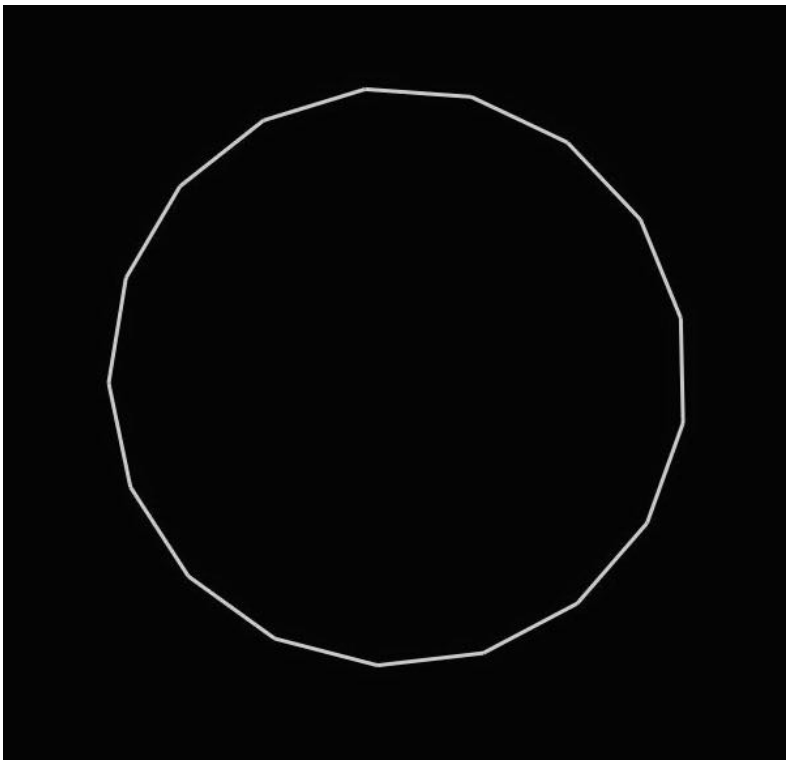

Diagram 6.4.
A heptadecagon produced either using Algorithm 6.2 or Algorithm 6.3.

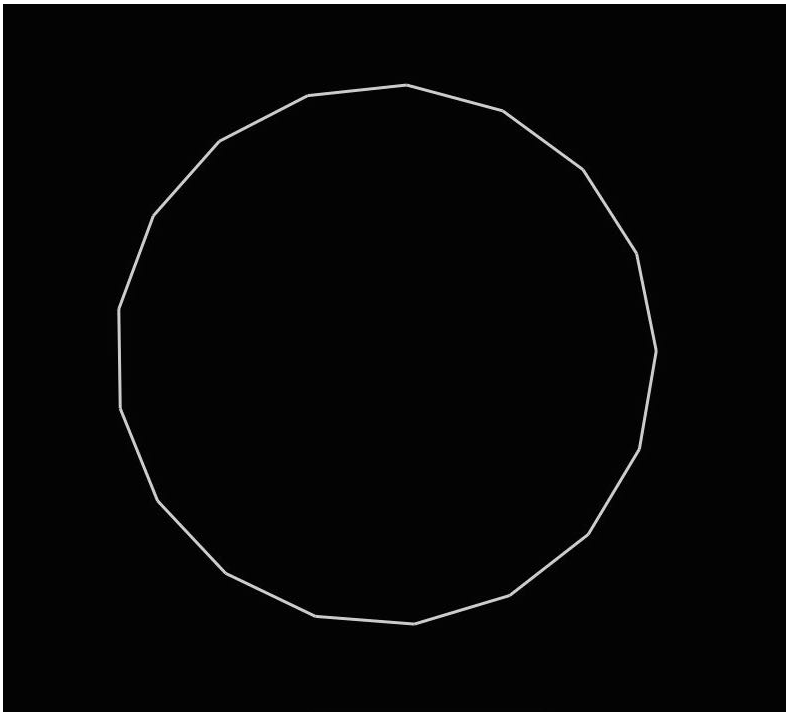

Diagram 6.5.
A heptadecagon produced using the other algorithm.

The second procedure also contains an idea that is far more general: it can be adapted to construct an approximately regular polygon of any number of sides. For instance, below is a polygon that approximates a regular heptagon, a shape for which Gauss' result implies no exact construction procedure can be given. The result is accurate to around four parts in a thousand (one side is slightly shorter than the other six).

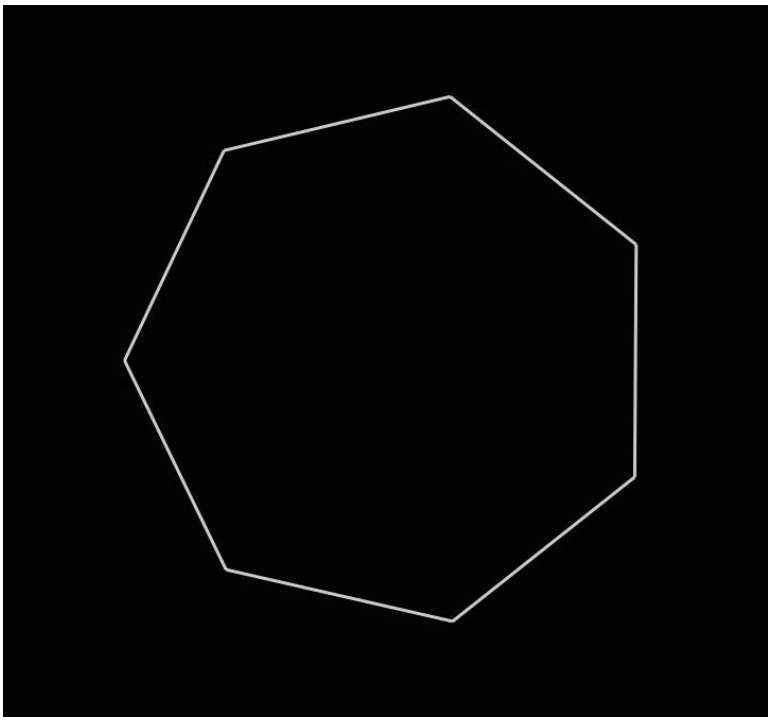

Diagram 6.6. A polygon that closely approximates a regular heptagon.

Though Algorithm 6.3 can be adapted to yield a polygon that is arbitrarily close to being regular, this will never achieve a completely regular construction. Yet we can now add to this an observation similar to that which motivated the central question attended to in this thesis. Even if we are very careful and use only the finest instruments, when we actually put Algorithm 6.2 into practice the resulting physical drawing will inevitably fall short of absolutely perfect regularity.

We define Implementation Errors for both algorithms as the expected practical deviation from an ideally constructed drawing, and the Internal Error of Algorithm 6.3 as the deviation of an ideal implementation from a truly regular polygon. We may assume for the sake of argument that there is a comprehensive physical argument establishing that when the algorithms are put into practice the sum of Internal and expected Implementation Errors for Algorithm 6.3 is smaller than the expected Implementation Error for Algorithm 6.2. This is in any case highly plausible as Algorithm 6.3 is much simpler. We have the following table:

	Ideal World	Real World
Algorithm 6.2	1	$1 - p_{imp}$
Algorithm 6.3	$1 - q_{int}$	$1 - q_{int} - q_{imp}$

Table 6.7, showing accuracies. p_{imp} and q_{imp} are the Implementation Errors, q_{int} is the Internal Error of Algorithm 6.3, and we may assume that $q_{int} + q_{imp} < p_{imp}$.

We can now engage in an analogue of the thought experiment conducted in the previous chapter. Suppose now that we need a drawing of a regular heptadecagon that is as precise as possible for some important purpose. Which method should we now choose? Would we not be *irrational* to prefer the method that is less accurate when actually implemented, even though it may work perfectly in theory?

The second algorithm may be superior in a sense, but Plato would rightly be more pleased with a geometer who had discovered the first. For this subject matter – within the context of Public Acceptance, where the Intellectual Virtues apply – the relevant comparison is the first column. In evaluating the two Algorithms *qua* pieces of mathematics we are concerned with their abstract content and not the empirical conditions under which they can be put into practice effectively. It is therefore clear that only Algorithm 6.2, but not Algorithm 6.3, gives an acceptable solution to the construction problem studied by Gauss.

Over the next four sections, I shall argue that Probabilistic Arguments are likewise unacceptable in the context of Public Acceptance. The cases are disanalogous in two ways, however. Firstly, in comparing the Trial Division and Rabin-Miller algorithms we do actually want to put them into practice, and for the results to gain Public Acceptance. We are not simply studying their formal properties in the abstract. Secondly, Algorithm 3.13 can get us a real prime number: in all likelihood, 66,997,813 is as genuine a prime as 48,140,819. If this is true, it will have provided us with an acceptable solution in the sense that the final answer given at the end is not deficient in any way. However, the Intellectual Virtues apply to entire discourses and not just to their conclusions.

6.ii. Probabilistic Arguments Reconsidered

> 'In testing primality of very large numbers chosen at random, the chance of stumbling upon a value that fools the Fermat test is less than the chance that cosmic radiation will cause the computer to make an error in carrying out a 'correct' algorithm. Considering an algorithm to be inadequate for the first reason but not for the second illustrates the difference between mathematics and engineering.' – *Hal Abelson and Gerald J. Sussman*[286]

[286] Hal Abelson and Gerald J. Sussman, *Structure and Interpretation of Computer Programs* (Massachusetts: MIT Press, 1996), section 1.2.

We now begin the assault on Probabilistic Arguments in earnest. Let us first reconstruct the underlying logic of our main probabilistic argument.

Major Premiss:	$\mathbb{P}(Algorithm\ 3.13\ outputs\ a\ composite\ ) < \varepsilon = 10^{-10,000}$
Minor Premiss:	$Algorithm\ 3.13\ outputted\ \ 66{,}998{,}713$
Conclusion:	$66{,}998{,}713\ is\ prime$

In Chapter 4, we raised a worry about whether a third party reading the argument can ever really have independent knowledge that the event expressed by the minor premiss actually occurred, though we ultimately did not consider this objection decisive. In Section 6.v below we raise further issues about the rule of inference appealed to here. Before that, however – in this section, and in the following three – we show that the two premisses rest upon assumptions that cannot be established in an acceptable way. The justification of these premisses involves the following theorem, restated here for convenience:

Theorem 6.8.

Let $n_1, n_2, n_3, \ldots$ be a sequence of candidates chosen randomly and independently from the set $\{2^{25}, 2^{25}+1, \ldots, 2^{26}-1\}$.

Suppose that for each candidate n_i in turn we randomly choose 16,613 elements from $\{1, \ldots, n_i - 1\}$ and test whether any of these are witnesses to the compositeness of n_i.

We output the first candidate for which no witnesses are found.

Then this procedure produces a prime number with a probability of at least $1 - \varepsilon$, where $\varepsilon = 10^{-10,000}$.

The statement of Theorem 6.8 here involves the concept of 'random choosing'. As was made clear our discussion of Univocality in Section 5.v, we ought to get clear on this phrase and ask what it really amounts to. Mathematical sense is made of the concept using the theory of discrete random variables initiated by Andreas Kolmogorov in the 1930's.[287] We will now briefly review the basics of his system, to gain a deeper insight into what is going on here.

Intuitively, the theory of random variables is best explained in the context of performing an experiment whose outcome is uncertain. Examples of experiments are tossing a coin and recording the number of accidents on a stretch of motorway in one hour. We now make some definitions, introducing only as much generality as is necessary for our purposes.

[287] Andrey Kolmogorov, *Foundations of the Theory of Probability* (New York: Chelsea, 1956). This is by far the most widely accepted approach, and though other approaches are available, this does not affect the argument.

By the term 'sample space', we denote the set of possible outcomes of the experiment. For instance, for the two experiments described in the previous paragraph, the sample spaces are $\{heads, tails\}$ and $\mathbb{N} = \{0,1,2,3 \dots\}$ respectively. Hence, although the members of Ω might be given numerically, this is not necessarily the case. For our purposes, we also need only consider cases where the sample space is countable.

Definition 6.10.

A *sample space* is a countable non-empty set Ω.

We also define 'events' as follows.

Definition 6.11.

An *event* $E \subseteq \Omega$ is a subset of the sample space.

For the second experiment, one event might be that an even number of accidents occurs during the one-hour interval observed. In rolling a normal die, two distinct events might be rolling an odd number and rolling a prime number. We say an event E has 'occurred' if the outcome of the experiment is contained in E.

Next we introduce a structure onto these events.

Definition 6.12.

Suppose $\mathcal{F}$ is a collection of events. Then $\mathcal{F}$ is called an *event space* iff the following three conditions are met:

1. $\Omega \in \mathcal{F}$
2. If $E \in F$ then $E^c \in \mathcal{F}$.
3. If $E_1, E_2, \dots, E_n \in \mathcal{F}$ then $\bigcup_{i=1}^{n} E_i \in \mathcal{F}$.

Such a collection of subsets of Ω is called a 'σ-algebra'. In rolling a die, where the sample space is $\Omega = \{1,2,3,4,5,6\}$, for most applications we can usefully take $\mathcal{F}$ to be the set of all $2^6 = 64$ subsets of Ω:

$$\mathcal{F} = \{\emptyset, \{1\}, \{2\}, \dots, \{2,3,4,5,6\}, \Omega\}.$$

In essence, a random variable is a way of assigning a numerical value to each outcome of the experiment. It is thus neither 'random', nor a variable: it is in fact a function. It must also be measurable: that is, for any real number x, the pre-image of x in Ω must always be a member of $\mathcal{F}$.

Definition 6.13.

A *random variable* is a function $X: \Omega \to \mathbb{R}$ such that $X^{-1}(x) \in \mathcal{F} \quad \forall x \in \mathbb{R}$, where $X^{-1}(x) = \{\omega \in \Omega | X(\omega) = x\}$.

For the first experiment described above, we might decide to stipulate $X(heads) = 1$ and $X(tails) = 0$. With the second we may simply take X as the identity function, as the outcomes of the experiment are already given as real numbers. Next we state conditions for ascribing probabilities to the events in an event space $\mathcal{F}$ in a way that is intuitively consistent.

Definition 6.14.

A *probability function* (or *probability measure*) is a function $\mathcal{P}: \mathcal{F} \to \mathbb{R}$ such that the following three conditions are met:

1. For all events $E \in \mathcal{F},\ \mathcal{P}(E) \geq 0$.
2. $\mathcal{P}(\Omega) = 1$.
3. For all $E, F \in \mathcal{F}$, if $E \cap F = \emptyset$ then $\mathcal{P}(E \cup F) = \mathcal{P}(E) + \mathcal{P}(F)$.

Definition 6.15.

A *probability space* is a triple $(\Omega, \mathcal{F}, \mathcal{P})$ subject to the above conditions.

Lastly, we are now able to define the probability mass function of the random variable X, using the preimage in Ω of elements of $Im(X) \subset \mathbb{R}$ and the probability function $\mathcal{P}$.

Definition 6.16.

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space. The *probability mass function* for a random variable $X: \Omega \to \mathbb{R}$ is a function $p_X: Im(X) \to \mathbb{R}$ given by

$$p_X(x) = \mathcal{P}(X^{-1}(x)) = \mathcal{P}(\{\omega \in \Omega | X(\omega) = x\})$$

We often write $\mathbb{P}(X = x)$ instead of $p_X(x)$.

As a familiar example, consider again the experiment of rolling a fair die together with the natural probability function $\mathcal{P}(\{1\}) = \mathcal{P}(\{2\}) = \cdots = \mathcal{P}(\{6\}) = \frac{1}{6}$, which extends uniquely to the function $\mathcal{P}(\mathrm{E}) = \frac{|E|}{6}$ for any $E \subseteq \Omega = \{1,2,3,4,5,6\}$. We can then obtain familiar results such as $\mathbb{P}(X\ is\ even) = \frac{3}{6} = \frac{1}{2}$. This is an example of a discrete uniform random variable – that is, one with a finite image set $Im(X) = \{x_1, x_2, \ldots, x_n\}$ and with $\mathbb{P}(X = x_i) = \frac{1}{n}$ for each $i, 1 \leq i \leq n$.

Using this concept of a discrete uniform random variable, we can now give a rigorous restatement of Theorem 6.8.

Theorem 6.17.

Let $(N_1, N_2, N_3, \dots)$ be an infinite sequence of independent discrete uniform random variables, each taking values in the set $\{2^{25}, 2^{25}+1, \dots, 2^{26}-1\}$.

Let $(A_{1,1}, A_{1,2}, \dots, A_{1,16613})$, $(A_{2,1}, A_{2,2}, A_{2,3}, \dots A_{2,16613})$, ... be an infinite set of sequences of random variables, each of length $16{,}613$, and where the $A_{i,j}$ are independent discrete uniform random variables on $\{1, \dots, N_i - 1\}$.

For each positive natural number i and each natural number j with $1 \leq j \leq 16{,}613$ we define a random variable $Y_{i,j}$ as follows:

$$Y_{i,j} = \begin{cases} 1 & if\, A_{i,j}\ is\ a\ witness\ to\ N_i \\ 0 & Otherwise \end{cases}$$

Now define another infinite sequence of random variables $Z_1, Z_2, \dots$ as follows:

$$Z_i = \sum_{j=1}^{16{,}613} Y_{i,j}^2$$

So that Z_i is zero if and only if no witnesses have been found for N_i.

Lastly, we put $r = \min\{i \ \ such\ that\ Z_i = 0\}$ and then $N = N_r$.

Then $\mathbb{P}(N\ is\ composite) < \varepsilon = 10^{-10{,}000}$.

This second statement of the theorem has no reliance on the mysterious and hitherto corporeal-sounding concept of 'random choosing'. Random variables are simply functions, and the theory of probability is the study of the formal properties of these functions. The theory of probability is thus a subfield as rigorous as any other branch of pure mathematics. However, for our argument to get running, we will need to venture beyond this formalism. Consider again the argument.

Major Premiss:	$\mathbb{P}(Algorithm\ 3.13\ outputs\ a\ composite) < \varepsilon = 10^{-10{,}000}$
Minor Premiss:	$Algorithm\ 3.13\ \ outputted\ \ 66{,}998{,}713$
Conclusion:	$66{,}998{,}713\ is\ prime$

As we have said, there is a further assumption that needs to be discharged here. Theorem 6.17 applies only when we have some sequences of independent discrete uniform random variables. In being able to successfully run algorithm 3.13, the compiler needs access to a random number generator: a device that supplies us with a stream of numbers. For the premisses to be true, we need it to be the case that the discrete uniform distribution gives the correct probability mass function for modeling the output of this random number generator. However, the meaning of this claim has not been fixed by anything we have said so far. We discuss this further in the next section, where it will become clear that it is not a mathematical question at all, but a question for the philosophy of probability.

6.iii. Interpreting Probability Statements

> 'It is unanimously agreed that statistics depends somehow on probability. But, as to what probability is and how it is connected with statistics, there has seldom been such complete disagreement and breakdown of communication since the Tower of Babel.' – *Leonard Savage* [288]

In the previous section, we saw that for the premisses of our Probabilistic Argument to be true we need it to be the case that the discrete uniform distribution gives the correct probability mass function for the output of our random number generator. The general question of when a given probability function is appropriate for modeling a concrete experiment is the philosophical problem of the interpretation of probability. But there is no settled account of how best to resolve this problem: only controversy and disagreement.

This disagreement concerns not only the details but is even about broadly what kind of account is needed. Is probability a logical concept, connecting bodies of evidence with hypotheses? Does it ascribe physical properties to experimental setups? Or is probability an epistemic notion – either determined by objective evidential conditions, or simply by subjective doxastic states? As we shall see shortly, all of these responses have been given.[289] There is thus a concern about whether Probabilistic Arguments are acceptable from the perspective of Univocality. This is a stark contrast to the mathematical theory of probability proper, which carefully avoids the issue here:

> 'The axioms of probability theory are set up so that abstract probabilities can be computed readily, but nothing is said about what probability really signifies, or how this concept can be applied meaningfully to the actual world.'[290]

Just as Hilbert's terms 'points, 'lines', and 'places' did not require interpretation within his axiomatic geometry, the mathematical theory of probability need never

---

[288] Leonard Savage, *The Foundations of Statistics* (New York: Dover, 1972), 2.
[289] This chapter's discussion of the various interpretations of probability is indebted to Alan Hájek, "Interpretations of Probability", *Stanford Encyclopedia of Philosophy*, accessed 22nd August 2015, http://plato.stanford.edu/entries/probability-interpret/
[290] Donald Knuth, *The Art of Computer Programming: Volume 2* (Addison-Wesley: Massachusetts, 1981), 142.

give guidelines for an appropriate choice of the underlying probability function $\mathcal{P}$. It simply leaves it as an arbitrary function, subject only to the conditions of Definition 6.14. This definition clearly satisfies both Abstractness and Univocality. Likewise, mathematically speaking Ω is simply defined as an arbitrary set, with no mention of experiments. The legitimacy of these definitions does not depend on any perceived correspondence with physical reality, though failure here might have meant that the theory was not useful.

The norm of Univocality tells us that in order for the premisses of a Probabilistic Argument to be acceptable from a mathematical point of view we must give a clear interpretation of the claim that the correct probability mass function for our random number generator is given by the discrete uniform distribution. However, it may be that this problem is not insuperable. Let us not overstate what is required: we need not resolve the general philosophical problem of the interpretation of probability in its entirety, but may simply stipulate a definition that will be adequate for this particular kind of experiment. However, I shall argue that none of the available interpretations will resolve the matter in a way that is acceptable from the perspective of both Abstractness and Explicitness.

Before proceeding with the argument, we first need to take note of another ambiguity in the problem. For our probability mass function may apply in two ways here. It can apply either directly to the hardware: the random number generating device itself, considered as a physical system. Alternatively, it can apply to the software this system runs: as a mathematical property of the algorithms used to generate the random numbers. Consequently, there will be two interpretations of the minor premiss as well. It could indicate that a particular corporeal event has occurred – that the physical random number generator did not find any numbers meeting the conditions for witnesshood for the candidate given. Or, it could also express a mathematical property of the number generating software: that it does not generate any such numbers.

The argument that no available interpretation is satisfactory will now proceed in three parts. In the rest of this section, we briefly discuss logical probability. In the next section, we consider the two remaining kinds of interpretations – physicalist and epistemic – as they apply directly to the hardware itself, showing that neither will be suitable. In Section 6.v, I give a parallel argument when the interpretations are taken to apply to the software running the algorithm. [291]

The theory of logical probability is so called because it regards probability as primarily applying to propositional inferences. According to this interpretation, whose noted proponents include Peirce, Keynes, Boole and Carnap, the probability associated with an inference from a given body of evidence E to an hypothesis H encapsulates the degree to which E provides evidential support for H. In evaluating probabilities we must thus consider a class of relevantly similar circumstances where inferences of this kind apply. This degree of support may be quantified as the proportion of instances when the inference yields us a true belief rather than a false one. Pierce elaborates:

[291] These sections will include discussion of the frequency interpretation, which is given both hardware and software-based formulations.

> 'The inference from the premise, A, to the conclusion, B, depends, as we have seen, on the guiding principle, that if a fact of the class A is true, a fact of the class B is true. The probability consists of the fraction whose numerator is the number of times in which both A and B are true, and whose denominator is the total number of times in which A is true, whether B is so or not.'[292]

It is not clear how such a definition can apply here. We are not looking to quantify relationships between hypotheses and bodies of evidence at all, but rather, to understand the conditions under which a random number generator should be modeled by a discrete uniform probability distribution on $\{1, \dots, n\}$. The only conceivable way this definition could apply is to let E be the claim that X is a random variable representing the output of our random number generator and let H_k be the hypothesis that $X = k$ for each $1 \leq k \leq n$. If we then take Peirce's proportional interpretation of the strength of evidential support, we arrive at simply another way of saying that the random number generator will output each element of $\{1, \dots, n\}$ with the same proportional frequency. This idea will be discussed later, under the guise of the frequency interpretation.

6.iv. Hardware-Based Approaches

In this section, we consider both epistemic and physicalist (including frequency) interpretations of probability as they apply to a random number generator *qua* physical piece of hardware. Crandall and Pomerance suggest one such method for generating random numbers using a physical process:

> 'Aim a microwave receiving dish at the remote heavens, listening to the black-body "fossil" radiation from the early cosmos, and digitize that signal to create a random bitstream.'[293]

In applying directly to a corporeal entity, any such interpretation must fail to meet Abstractness. When the premisses are set out in full then our argument will be seen to contain propositions that essentially refer to a physical system. However, because one might take the view that it is worthwhile to sacrifice Abstractness in this instance in return for the other benefits probabilistic algorithms can bring, we will also discuss each of the available interpretations in more detail. We will see that each is attended with its own peculiar problems when employed in this context. I further argue that none of these interpretations can be applied in a way that is acceptable from the perspective of Explicitness.

Let us begin by discussing the classical theory of probability. This interpretation was perhaps first anticipated by Cardano in his *Liber de Ludo Aleae*, and discussed by Leibniz, Bernoulli, de Moivre. Another noted source is the Fermat-Pascal correspondence – especially their discussion of games of chance. This was occasioned by a question posed by the Chevalier de Meré, who was interested in

[292] Charles Sanders Peirce, "The Red and the Black", in *The World of Mathematics*, ed. James Newman (New York: Dover, 1956), 1336.
[293] Crandall and Pomerance, *Prime Numbers*, 361.

the problem of how to split winnings fairly when such games were interrupted before their completion.[294] But arguably the most influential statement of the theory is due to Laplace:

> 'The theory of Chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favourable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability, which is thus simply a fraction whose numerator is the number of favourable cases and whose denominator is the number of all the cases possible.'[295]

The idea, then, is to apportion probabilities equally amongst cases between which the evidence is equally balanced, or between which our total information is indifferent. Consider again the experiment of throwing a die that is made from uniform material. There is no reason to think the outcome will be any particular number rather than another, and so we assign each outcome equal probability. Similar examples are common in probability and statistics textbooks. The following example is fictitious but may be taken as typical:

> Suppose I have a bag with 8 balls, each identical except that 5 are red and 3 are yellow. I put my hand in and pull a ball out without looking. What are the chances I pull out a red ball?

Laplace gives his definition immediately after his even more famous statement of the principle of causal determinism. So accordingly, if we know to a sufficient degree the initial conditions of a particular experiment, we may know the outcome with certainty: 'probability is relative, in part to this ignorance, in part to our knowledge.'[296] It is clear then that the classical interpretation of probability is indeed epistemic in character. Probabilistic statements do not give objective descriptions of external states of affairs, but rather, indicate something about the evidential state of those who utter them.

As we saw with epistemic probabilities more generally in Chapter 3, this interpretation is clearly problematic from the perspective of the Practical Virtues. Most of the time enquirers will have different epistemic states, damaging Consensus, and the probabilities may be updated over time, damaging Permanence. So this interpretation is unsuitable for bringing Probabilistic Arguments into compliance with the demands of Univocality within the context of Public Acceptance for these same reasons. However, we will also point out a further problem with epistemic interpretations in this context.

---

[294] Prakash Gorroochurn, "Some Laws and Problems of Classical Probability and How Cardano Anticipated Them", *Chance* 25 (2012): 13–20. See also Stephen Fienberg, "A Brief History of Statistics in Three and One-half Chapters: A Review Essay", *Statistical Science* 7 (1992): 208–225 and James Franklin, *The Science of Conjecture: Evidence and Probability before Pascal* (Maryland: The Johns Hopkins University Press, 2001).

[295] Pierre Simon de Laplace, *Concerning Probability*, in *The World of Mathematics*, *Vol 1*, ed. James Newman (New York: Dover, 1956), 1327.

[296] Ibid., 1326.

Suppose we take a newspaper and extract all the numbers found written anywhere within it, forming them into a list. We then pick a number at random from our list and record its first non-zero digit. The sample space for this experiment is $\{1,2,\ldots,9\}$. Supposing we know nothing else about this situation. Then we have no reason to believe that the outcome will be any particular integer rather than another, and so the classical theory of probability suggests we apply a uniform distribution. However, doing so would be ill advised here, because – somewhat surprisingly – the true probability has been empirically determined to be as follows:

$$p(d) = \log_{10}(1 + \frac{1}{d})$$

This is known as 'Benford's Law', and has a number of partial and overlapping explanations.[297] In establishing this result, we take empirical observation to trump the application of the classical theory in this instance. This is a serious problem for the classical interpretation: it could be that our evidence is indifferent between every possible outcome only because we lack important knowledge about the experimental setup. It is not enough to plead ignorance: if the true probabilities are not in fact uniform, this may affect the running of our algorithm. We need to have positive knowledge of the correct assignment.

Next we discuss the subjective interpretation of probability, which is largely due to Frank P. Ramsey.[298] Ramsey was sceptical about the existence of objective logical relations between pieces of evidence and hypotheses, and hence the determinateness of the assignments suggested by the logical interpretation discussed in the previous section. He argued instead that probability is the 'logic of partial belief'; the degree of credence experienced by agents – specified to be in some sense rational, to avoid consequences of the kinds of statistical biases discussion in Section 3.vi – in the face of uncertainty and in response to particular situations or bodies of evidence. This conception of probability thus lives up to its name: 'Probabilistic reasoning – always to be understood as subjective – merely stems from our being uncertain about something.'[299]

Clearly, the same problem just raised appears again. The naïve approach of fitting a uniform distribution to the newspaper data is not indicative of a failure of rationality *per se*, but merely a lack of empirical knowledge. But as well as this, there are also further problems with both Abstractness and Univocality that will prove to be insurmountable. To assert that our random number generator is such that rational agents would give equal assent to the propositions that it will output each number in the range automatically brings into our Probabilistic Arguments the concepts 'rational', 'agent', and 'belief'. These are notions that find their application not to *abstracta* but to complex intentional systems such as ourselves, and which cannot be formulated sufficiently clearly or precisely for the purposes

[297] Arno Berge and Theodore Hill, *An Introduction to Benford's Law* (Princeton: Princeton University Press, 2015).
[298] Frank Ramsey, "Truth and Probability", in *Foundations of Mathematics and Other Logical Essays* (Abingdon: Routledge, 2006).
[299] Bruno de Finetti, *Theory of Probability: Vol 1* (New York: Wiley, 1990), preface.

of mathematical theorising. For instance, 'belief' has typically been understood here in terms of ideal betting or exchange behaviour:

> 'Let us suppose that an individual is obliged to evaluate the rate p at which he would be ready to exchange the possession of an arbitrary sum S (positive or negative) dependent on the occurrence of a given event E, for the possession of the sum pS; we will say by definition that this number p is the measure of the degree of probability attributed by the individual considered to the event E, or, more simply, that p is the probability of E (according to the individual considered; this specification can be implicit if there is no ambiguity).'[300]

Moreover, in being so explicitly subjective the concept thus elucidated is again clearly unsuitable for our purposes here because there is explicit provision for the possibility that different agents disagree in assignments. Again, licensing propositions expressing subjective probability statements in the context of Public Acceptance would therefore damage both Consensus and Permanence.

Let us now see if physicalist accounts of probability can avoid these problems with modeling our random number-generating hardware. Consider first the best systems approach of David Lewis, which builds on his account of the laws of nature.[301] Lewis regards the laws of nature as the theoretical framework that gives a description of the universe that is the best possible in terms of theoretical virtues such as simplicity and strength (i.e. predictive power). If we include probabilistic theories such as quantum mechanics, then we can introduce another theoretical virtue called 'fit'. A framework has more fit than another if the actual history of the universe is asserted to be more probable within that framework. We then ascribe probabilities to particular experiments just in case the ultimate laws of nature describe them in a probabilistic manner.

Again, whilst interesting, Lewis' efforts will not be of much help to us here. For although there is strong evidence for quantum mechanics, it is still not an attractive idea to hold mathematics hostage to empirical enquiry. We are currently hardly in a position to give a sufficiently confident answer to the ultimate theory of everything, and yet it appears that it would be necessary for us to do so before we could sure that we were ascribing the appropriate probability function to our random number generator.

A second physicalist account is the propensity interpretation. This framework has received perhaps its fullest development in the hands of Karl Popper,[302] though it was anticipated by Peirce in 1910.[303] Like Lewis' best systems theory, it is

---

[300] Bruno de Finetti, "Foresight: Its Logical Laws, Its Subjective Sources", in *Studies in Subjective Probability*, ed. H. E. Kyburg, Jr. and H. E. Smokler (New York: Robert E. Krieger Publishing Company, 1980), 62.
[301] David Lewis, "Humean Supervenience Debugged", *Mind* 103 (1994): 473-490.
[302] Karl Popper, "The Propensity Interpretation of Probability", *The British Journal for the Philosophy of Science* 10 (1959), 25-42.
[303] Charles Sanders Peirce, "Note on the Doctrine of Chances", in *Collected Papers of Charles S. Peirce: Vol II*, ed. Charles Harteshorne, Paul Weiss, and Arthur Burks (Massachusetts: Harvard University Press, 1931-1958), section 408.

intended to constitute an objective description of the physical world. This time we ascribe to the random number generator (or perhaps to the entire physical set-up) a certain kind of propensity, or disposition, to yield each possible outcome – either in a single case or in the long run.

The approach seems promising though we might again worry that the notion of propensity cannot be formulated with sufficient precision to comply with Univocality. Consider in particular Pierce's interpretation of assigning the usual discrete uniform probability distribution to a die:

> 'The statement means that the die has a certain 'would-be'; and to say that the die has a 'would-be' is to say that it has a property, quite analogous to any habit that a man might have.'[304]

Later in the section we will enquire how quantitative knowledge of these propensities might be achieved. But first let us discuss one final physicalist interpretation of probability.

The final theory we shall consider is the frequency interpretation, developed in an 1876 work by Venn.[305] Other names associated with the frequency approach are Reichenbach, von Mises, and Fisher. In the next section, we will see that it can be given a software interpretation: here we again understand it empirically, as applying to the hardware used to generate the numbers. In its more tenable versions, it states that the probability function $\mathcal{P}: \mathcal{F} \to \mathbb{R}_0^+$ is to assign the numerical value q to the probability of the event $E \in \mathcal{F}$ if and only if q is the long-term limiting proportion into which outcomes of the experiment would fall into the set $E \in \Omega$ if the experiment were repeated infinitely many times.

The frequency approach is highly popular with both statisticians and natural scientists, so let us examine this definition in more detail. Suppose that we have an actual experimental situation with a set of possible outcomes Ω, and whose outcome is modeled by a single random variable $X: \longrightarrow \mathbb{R}$. Imagine now that we were to repeat the experiment infinitely many times 'under similar conditions' and obtain successive outcomes $\omega_1, \omega_2, \ldots$. We could then define sequences $\langle x_i \rangle$ and $\langle y_i \rangle$ by setting $x_i = X(\omega_i)$ for all i and then $y_i = \left|\{j \mid 1 \leq j \leq i \ \ and \ x_j = k\}\right|$, so that y_i is the number of times that X takes the value k in the first i repetitions of the experiment. This interpretation of probability says that we should define $\mathbb{P}(X = k)$ to be the number p if and only if $\lim_{n \to \infty} \frac{y_n}{n} = p$.

As the scare-quotes suggest, there are problems in making sense of the phrase 'under similar conditions' here. In many cases, if the initial conditions are sufficiently similar then we will always get the same outcome each time: for instance, mechanical coin flipping machines exist which can reliably give a predetermined outcome.[306] In general we will not be able to give clear necessary and sufficient conditions but only indicate roughly what is meant here, perhaps

[304] Peirce, "Note on the Doctrine of Chances".
[305] John Venn, *The Logic of Chance* (New York: Chelsea Publishing Co., 1962).
[306] See also Joseph B. Keller, "The Probability of Heads", *The American Mathematical Monthly* 93 (1986): 191-197.

through the use of examples. As we shall see presently, there are problems with how we would ever know this interpretation to apply. Indeed, we may also worry that in practice it will never apply, as the Cauchy/Weierstrass limit formulation is so exacting that even a tiny asymmetry in our die will prevent it from being literally true.[307] To contrast the character of an application of this definition with a piece of genuine mathematics, we give a short discussion of a related theorem.

The law of large numbers applies when we have a countably infinite sequence $\langle X_i \rangle$ of random variables defined on a sample space Ω (that is, a sequence of functions $X_i : \Omega \longrightarrow \mathbb{R}$) which are pairwise independent and such that each has the same probability mass function with mean μ and variance σ^2. This time, for each ω in Ω we define an associated sequence $\langle x_i^\omega \rangle$ with $x_i^\omega = X_i(\omega)$ for all i, and a second sequence $\langle \bar{x}_j^\omega \rangle$ as the arithmetic mean of the first j terms of the first sequence, i.e. $\bar{x}_j^\omega = \frac{1}{j}\left(x_1^\omega + \cdots + x_j^\omega\right)$.

If the variance is finite, the strong law of random numbers says that however the underlying probability function $\mathcal{P}$ is defined we always have that:

$$\mathcal{P}\left(\left\{\omega \mid \lim_{n\to\infty} \bar{x}_n^\omega = \mu\right\}\right) = 1$$

Likewise, the weak law of large numbers, which does not require the variance to be finite, says that

$$\forall \varepsilon > 0, \lim_{n\to\infty} \mathcal{P}(\{\omega \mid \ \mu - \varepsilon < \bar{x}_n^\omega < \mu + \varepsilon\}) = 1$$

There are also various generalisations that relax other conditions. Unlike a concrete application of the frequency definition, these results can be explicitly proved like any other piece of genuine mathematics, such as the Pythagorean Theorem. Being authentic theorems, their veracity is unquestioned.

Note also that for these two mathematical results there is a single experiment with outcome ω that determines the value of all the X_i, and so when we state each theorem we can make use of the usual multiplicative definition of pairwise statistical independence:

$$\mathbb{P}\left(X_{n_1} = x_1, .., X_{n_k} = x_k\right) = \mathbb{P}\left(X_{n_1} = x_1\right) ... \mathbb{P}\left(X_{n_k} = x_k\right)$$

where the n_i are distinct.

But for the frequentist interpretation, though we likewise need some condition saying that successive repetitions of the experiment do not affect each other, this mathematical formulation will not be available because there is no larger probability function applying to joint outcomes of successive repetitions of the

[307] For further criticisms of the frequency interpretation, see Alan Hájek, "Fifteen Arguments Against Hypothetical Frequentism", *Erkenntnis* 70 (2009): 211-235.

experiment. We must therefore rely on the intuitive concept of causal independence, which though having an important place in empirical science is less precise than its mathematical counterpart.

These physicalist interpretations of probability each give empirical conditions for a discrete uniform distribution to apply to a given experiment. Let us now think about how we might secure knowledge that they are met, returning to the simple case of throwing a die. I have now in front of me a die on the table. Suppose now I am to roll it. Regardless of which interpretation of probability we advocate, it is attractive to believe that I know that the chances of rolling any of the six possible numbers are identically one in six. But how exactly do I know that this is the appropriate probability distribution for the experiment? How do we determine what the die's propensity to roll each number is, or the frequency with which each outcome will arise if we were to roll it over and over again, infinitely many times?

One important line of thought here is the symmetry of the die: its uniform physical construction suggests the discrete uniform distribution. There are of course people who take this kind of question far more seriously: how do Last Vegas hotels ensure that a batch of dice intended for use in one of their casinos are fair? One would imagine there are a number of methods here: attention to the manufacturing process, perhaps supplemented with empirical tests involving statistical frequency analysis, whereby samples of dice are tested to see if they deviate from expected behaviour when thrown a certain number of times.

No physical die is ever perfectly symmetrical on a sufficiently small scale, however: these enquiries are only approximate. If we had sophisticated enough measuring equipment, we would perhaps realise that we should be giving probabilities differing slightly from $1/6$. The actual die I have here is made from wood: how will the direction of the grain affect how it bounces? The numbers are marked with painted depressions: how does this affect its weight distribution? We are not sure exactly how it will be thrown either – so it is not clear what impact these inequities might have. But perhaps we should be less sure that a uniform distribution really is the true answer. Certainly it will not be possible to fully establish this claim through a chain of reasoning that is acceptable from the standpoint of Explicitness. Indeed, research suggests that a hand-flipped coin is biased towards returning to its initial orientation:[308] the probability it comes up the same way is around 0.51.

This example makes clear that our justification for assigning the uniform distribution to a random number generator *qua* physical piece of hardware cannot comply with Explicitness. Thus concludes our discussion of the hardware approach. For all the interpretations of probability available here, the conditions given will clearly fail Abstractness. Moreover, in practice a justification for applying a uniform distribution cannot be given without the violation of Explicitness. Each approach was also shown to have additional problems in this context. In the next section, we consider the application of probability directly to the software being run on our machines when implementing our algorithms.

[308] Persi Diaconis, Susan Holmes, and Richard Montgomery. "Dynamical Bias in the Coin Toss", *SIAM Review* 49 (2007): 211-235.

6.v. Software-Based Approaches

The arguments of the last chapter show that we cannot regard probability statements as referring directly to a random number generator *qua* physical piece of hardware if our four Intellectual Virtues are to be adhered to. In this section we apply our distributions directly to the underlying software, thus avoiding a conflict with Abstractness. Nevertheless, we shall see that no argument capable of being broken down into clear and simple steps can be given for the conclusion that the discrete uniform probability distribution correctly models our random number generator. Probabilistic Arguments therefore inevitably fail to meet with the Intellectual Virtue of Explicitness.

When so-called random numbers are supplied by a computer, typically what happens in practice is that the compiler uses a pseudorandom number generator that is in fact entirely deterministic, though it is thought hard to predict its behaviour in practice (again this often cannot be proved). Such pseudorandom number generators take as input a 'seed' value – usually taken from the computer's internal clock – and repeatedly apply a complicated and (apparently) difficult-to-invert number-theoretic function to it, producing a string of numbers recursively.

One way of modeling these generators is to regard the seed value ω as the outcome of the physical experiment of running the program. Once the seed value is determined then so is our sequence, and ultimately the output too. So we can regard the generator as a representing a function $F: \Omega \to \mathbb{N}$. But this would be a return to the hardware approach because we would be modeling a physical experiment. So instead we focus on the sequences generated and enquire under what conditions we can consider them to be random. Commenting on this situation, von Neumann wrote in 1951 that 'Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin.'[309] The use of such pseudorandom sequences is common in practice, however, and this is the approach taken. So let us first get more acquainted with pseudorandom sequences.

We begin with a highly simple example: a linear congruential generator. To construct a sequence of random numbers between 0 and $m-1$, we begin with a seed. In this case, I'll take the current time in minutes – it is 14.51, so 891. We then compute successive terms using a recurrence relation as follows:

$$x_{n+1} \equiv ax_n + b \bmod m$$

Arbitrarily picking $a = 1000$ and $b = 314$, we use this to generate a sequence of possible witnesses to the primality of 1729:

$$891, 879, 982, 242, 254, 151, 891, 879, \ldots$$

[309] Quoted in Knuth, *The Art of Computer Programming: Vol 2*, 3.

The sequence is periodic with a period of only 6! We would have been fortunate here because 891 is a witness:

$$891^{1728} = 1 \, mod \, 1729$$
$$891^{864} = 1 \, mod \, 1729$$
$$891^{432} = 1 \, mod \, 1729$$
$$891^{216} = 1 \, mod \, 1729$$
$$891^{108} = 1 \, mod \, 1729$$
$$891^{54} = 1065 \, mod \, 1729.$$

However, it is clear that this particular method is inadequate for our general purposes. Any such sequence must also be periodic if it is generated recursively: eventually we must arrive at a term previously encountered, as there are only finitely many terms to choose from. Past a certain point it would then not matter how many IMC tests we perform in total: only the number of distinct elements in the sequence will be important. If this number is more than $(n-1)/4$, then the test is guaranteed to be deductively conclusive. But it might be much less than this, as the example just given shows.

Knuth writes 'Many random number generators in use today are not very good … It is not easy to invent a foolproof source of random numbers.'[310] He then goes on to describe an algorithm he once invented that unexpectedly turns out to yield a sequence that eventually becomes constant for some seed values. He comments: 'The moral of this story is that *random numbers should not be generated with a method chosen at random*. Some theory should be used.'[311] Again, what is required here is competence – not the rational management of ignorance: we need to have positive knowledge that the sequences we are using will work well in practice. So as before, the subjectivist approaches are no good here. Let us see if the frequency approach fares any better.

We reduce the problem to constructing binary sequences as follows. Suppose we want to pick an integer from the set $\{a, a+1, \dots, b\}$ where $2^m \leq a < b < 2^n$. We generate $n-1$ binary bits and see if the resulting number lies in $\{a, \dots, b\}$. If not, we simply discard it and look for another number.

Suppose now that our generator gives us the alternating sequence 0,1,0,1,0,1,0, ...

Intuitively, this sequence does not seem 'random' at all. Yet it is clear that it meets the conditions stipulated by the frequency definition: the proportion of terms equal to 0 or 1 tends to $1/2$. If we used this for Algorithm 3.13 it would always select the same even number $10101010 \dots 101 = 2^{25} + 2^{23} + \cdots + 2$ in binary, and for Rabin-Miller applications more generally would only ever check a single witness.

So, when is a pseudorandom sequence sufficiently 'random' to produce the kind of results we need in a Rabin-Miller test? What general conditions can we give? As Knuth explains, 'A quantitative definition is needed. It is undesirable to talk about

[310] Ibid., 4.
[311] Ibid., 5.

concepts that we do not really understand, especially since many apparently paradoxical statements can be made about random numbers.'[312] He goes on to consider two definitions by earlier authors:

> 'D. H. Lehmer (1951): "A random sequence is a vague notion embodying the idea of a sequence in which each term is unpredictable to the uninitiated and whose digits pass a certain number of tests, traditional with statisticians and depending somewhat on the uses to which the sequence is to be put."
>
> J. N. Franklin (1962): "The sequence (1) is random if it has every property that is shared by all infinite sequences of independent samples of random variables from the uniform distribution."'[313]

Concluding that neither definition is adequate, Knuth then writes:

> 'What we really want is a relatively short list of mathematical properties, each of which is satisfied by our intuitive notion of a random sequence; furthermore, the list is to be complete enough so that we are willing to agree that any sequence satisfying these properties is "random."'[314]

As Lehmer suggests, in practical applications theoreticians tend to make use of a number of distinct tests to check if a given random number generator is appropriate in a given context. Knuth explains:

> 'Two kinds of tests are distinguished: *empirical tests*, for which the computer manipulates groups of numbers of the sequence and evaluates certain statistics; and *theoretical tests*, for which we establish characteristics of the sequence by using number-theoretic methods based on the recurrence rule used to form the sequence.' [315]

Further to the frequency test, Knuth discusses several other such tests: the serial correlation test, the run test, the gap test, the maximum-of-t test, and the collision test amongst others. He goes on:

> 'The reader probably wonders, "*Why are there so many tests*?" … The need for making several tests has been amply documented. It has been recorded, for example, that some numbers generated by a variant of the middle-square method have passed the frequency test, gap test, and poker test, yet flunked the serial test. Linear congruential sequences with small multipliers have been known to pass many tests, yet fail on the run test because there are too few runs of length one. The maximum-of-t test has also been used to ferret out some bad generators that otherwise seemed to perform respectably.'[316]

When a new kind of algorithm is invented, more testing will need to be done to see whether a given generator gives good results when paired with it, and new kinds of

[312] Ibid., 142.
[313] Ibid., 142.
[314] Ibid., 142.
[315] Ibid., 39.
[316] Ibid., 73.

applications may even necessitate entirely new tests. Moreover, problems with generators have been quite frequent in the past:

> 'It seems that just as often as a new random-number generator is developed, so, too, is some older scheme shown to be nonrandom enough to be, say, "insecure," or yield misleading results in Monte Carlo simulations.'[317]

Consider also Knuth's lamentations on this topic:

> 'For more than a decade, the most common random number generators in daily use were seriously deficient'.[318]

Clearly, heuristic tests such as these can never be entirely conclusive: we are adopting something like an engineering perspective here, relying on a mixture of experience, empirical observation and intuition as well as calculation. The field of pseudorandom number generation focuses on developing methods that work well in practical applications in cryptography and other applied computer sciences, rather than rigorous derivation. In particular, the suitability of a particular pseudorandom number generator for a given application can never be established in such a way as to comply with the demands of Explicitness.

Let us again compare the situation to a similar one within mathematics proper. It is often of interest know whether a given series such as

$$\sum_{n=1}^{\infty} \frac{1}{n} \qquad \text{or} \qquad \sum_{n=1}^{\infty} \frac{1}{n^2}$$

converges or not. In this case, the former (the harmonic series) diverges, whereas the latter converges[319] to $\frac{\pi^2}{6}$, as shown by Euler. In approaching problems of this nature we also we have a number of tests that we can apply, where the spheres of application of each test overlap, and no test is universally applicable. These include the ratio test, the root test, the comparison test, the alternating sequence test, the integral test, and Abel and Dirichlet's tests. Yet though such tests can be inconclusive, when they do yield a definite verdict this always constitutes a rigorous proof that the sequence is convergent. Any further testing is then otiose.

In this section, we have seen that software approaches fare no better than hardware approaches in giving an account of Probabilistic Arguments that is acceptable from the perspective of Univocality. We have seen that there is no determinate, objective, operational criterion to tell us when a given pseudorandom sequence may be modeled with a discrete uniform distribution for the purposes of a given Monte Carlo method. Practitioners instead rely on a series of heuristic tests whose application does not meet the demands of Explicitness. In the next section, we

---

[317] Crandall and Pomerance, *Prime Numbers*, 361.

[318] Knuth, *The Art of Computer Programming: Vol 2,* 4.

[319] A series $\sum_{n=1}^{\infty} a_n$ is said to converge if and only if the sequence $\langle s_m \rangle$ of partial sums $s_m = \sum_{n=1}^{m} a_m$ converges.

discuss the nature of inference within Probabilistic Arguments, showing that it precludes them from being Formalizable.

6.vi. Probabilistic Inference

In this section, I consider Probabilistic Arguments from the perspective of Formalizability, the Intellectual Virtue articulated in Section 5.vi. We will see that as things stand Probabilistic Arguments cannot be formalised within a standard formal system. Moreover, attempts to augment these formal systems to accommodate Probabilistic Arguments will also be shown to be problematic. Hence, we will show that mathematicians have further reasons to reject these arguments in the context of Public Acceptance.

Consider again the form of our Probabilistic Argument from Chapter 3, which I will again repeat here for convenience.

Major Premiss:	$\mathbb{P}(Algorithm\ 3.13\ outputs\ a\ composite) < \varepsilon = 10^{-10,000}$
Minor Premiss:	$Algorithm\ 3.13\ \ outputted\ \ 66{,}998{,}713$
Conclusion:	$66{,}998{,}713\ is\ prime$

It is clear that the inference from the premisses to the conclusion cannot be constructed within the kinds of standard formal system we have discussed. For the inferences of such formal systems are always truth preserving. But for arguments of this kind – based on Monte Carlo procedures – it is possible for the premisses to be true and the conclusion false. So in order to formalise the argument, we need to add a new rule of inference to the standard formal systems within which Publicly Acceptable mathematics is to be formalised; one that explicitly licenses these kinds of inferences.

In an old joke attributed to American philosopher Morris Cohen, logic texts are divided into two parts. In the first part, concerning deductive logic, deductive fallacies – wherein conclusions are drawn that do not follow logically from the premises – are guarded against. In the second part, concerning inductive logic, they are endorsed.[320] Needless to say, many mathematicians will take the first perspective here and curtly dismiss such an inference rule as endorsing a fallacy. We will press the matter a little further, however, as inferences of this sort are clearly common in natural science as well as everyday life. Nevertheless, I shall give two reasons against introducing a new rule of inference of this sort into the logic in which Publicly Accepted mathematics is to be formalised. Consider first the following argument schema.

[320] Vincenzo Crupi, "Confirmation", *Stanford Encyclopedia of Philosophy*, introduction, accessed 17th August 2015, http://plato.stanford.edu/entries/confirmation/

Major Premiss:	$\mathbb{P}(Algorithm\ 3.13\ outputs\ a\ composite\ ) < \varepsilon = 10^{-10,000}$
Minor Premiss 1:	$Algorithm\ 3.13\ outputted\ \ N$
Minor Premiss 2:	$N\ is\ composite$
Conclusion:	$N\ is\ prime$

This argument represents a situation where we have been incredibly unlucky in having Algorithm 3.13 fail us by outputting a composite number N, which we also happened to already know was composite (perhaps it ends in an even digit, say). Clearly, in this situation and knowing these premisses we would not be justified in coming to believe the conclusion. But if the second minor premiss were omitted, the argument would be acceptable within our modified formal systems equipped with a rule permitting probabilistic inferences.

This shows that the logic of any formal system that could express Probabilistic Arguments would not be monotonic. A logic is monotonic if whenever a proposition φ can be deduced from a set of propositions Γ, it can also be deduced from any set of propositions Γ' such that $\Gamma \subseteq \Gamma'$. But choosing a non-monotonic logic as the basis for formalizing Publicly Accepted mathematics is unacceptable, on pains of sacrificing Permanence. If we were to do so, later practitioners might come to know new propositions that invalidate entire classes of earlier inferences. Mathematics could therefore no longer be a cumulative hierarchy of theorems that grows over time.

Another issue is that unlike rules of inference known to be truth-preserving, probabilistic rules of inference cannot be used arbitrarily many times within the same argument. To illustrate why not, consider the following iterated iterated iterated Monte Carlo Algorithm.

Algorithm 6.20

Run Algorithm 3.13 repeatedly until it has generated $10^{1,000,000}$ integers.

Assert the proposition expressing the claim that all of these integers (which will not all be distinct) are prime.

Now, in this case we have a conclusion that is acceptable within our new modified formal system, as we may assume there are no unstated premisses that block the inference. But this application of the rule of inference is again clearly illegitimate: we have relied on it so many times that now it is actually overwhelming *un*likely that the conclusion will be true.

Again, things would be even worse if we gave a probabilistic proof by induction, where the inductive steps were justified only by separate appeals to the ε-rule in each instance, so that it was in fact relied upon infinitely many times:

Premiss 1:	$S(1)$
Premiss 2:	$\forall k,\ S(k) \overset{\varepsilon}{\rightarrow} S(k+1)$
Conclusion:	$S(n)\ \forall n \in \mathbb{N}$

Why is it that these are illegitimate invocations of the probabilistic inference rule when its use in determining that 66,998,713 is prime was not? And how many times can we use it, exactly? It is clear that we need an explicit condition here. For if it is left to individual judgement in particular cases, then this will introduce the possibility of disagreement, damaging Consensus. Moreover, if such a side-condition has not been given then our new formal systems will be problematic from the perspective of proof theory, where we are often interested in proving results about the entire class of acceptable derivations within a formal system. But if we have not specified how many times the probabilistic rule of inference can be appealed to, then the class of acceptable derivations becomes vague. Univocality is thus sacrificed, and the rigorous techniques of proof theory cannot be applied.

One idea for overcoming this problem is to introduce a global error constraint that applies to derivations as a whole. Suppose a derivation contains a finite number of probabilistic inferences with associated errors $E_1, E_2, \ldots, E_k$. As well as the requirement that each E_k is less than ε, we can also give a side condition that the total error E is less than ε. That is:

$$E = 1 - \prod_{i=1}^{k}(1 - E_i) < \varepsilon$$

Three further problems arise here. The first is of a practical nature. Because in practice proof presentations rely on subsidiary results, the total error associated with the derivations establishing these must always be taken into account and kept track of. Secondly, meeting this side-condition is a holistic property that requires us to look at a derivation in its entirety, whereas the techniques of proof theory tend to operate by looking at smaller chunks of an argument and using recursion. Lastly, there is also again concern about whether the specification of an acceptable derivation is sufficiently precise to meet Univocality. In the previous two sections we have seen that there may be disagreements about what the correct probability of error is. There is no definite procedure available for determining the total probability of error exactly: indeed, throughout the thesis we have been working only with upper bounds here.

6.vii. Conclusion

In this chapter, we have presented semantic, metaphysical, epistemic and logical considerations in favour of the conservative view that Probabilistic Arguments should not be given as justification in the context of Public Acceptance.

Univocality requires us to give a clear meaning to the claim that our random number generator should be modeled using a discrete uniform random variable. Doing this requires us to choose between the various interpretations of probability postulated by theorists hitherto, applied in either a hardware or a software sense. Every such application met with insuperable problems with Explicitness, and hardware approaches inevitability encounter problems with Abstractness too. Lastly, we saw in the final section that Probabilistic Arguments cannot be satisfactorily formalised.

We therefore conclude that discourses expressing Probabilistic Arguments cannot be brought into compliance with the four Intellectual Virtues of mathematical enquiry.

7. Should Mathematicians Play Dice?

> 'Wherever there is any hope of salvage, we will carefully investigate fruitful definitions and deductive methods. We will nurse them, strengthen them, and make them useful. No one shall drive us out of the paradise which Cantor has created for us.' – *David Hilbert*[321]

7.i. Concluding Summary

Let us review the argument given so far.

In Chapter 1 we articulated the four Practical Virtues of mathematical enquiry: Permanence, Reliability, Consensus and Autonomy. We argued that these were highly valuable to mathematicians, and also that insistence on proof prior to publication plays a central role in maintaining them.

In Section 1.vi, we then pointed out that these standards are now in danger of going into decline, due to the length and complexity of many proofs in contemporary mathematics, and the increasing reliance on computers.

In light of these problems, in Chapter 2 we investigated the use of deductive techniques in the context of justification. We concluded that only clearly defined methods that can be associated with a determinate, quantitative evidential value would be suitable for supporting the Public Acceptance of new results, on pain of a much more severe deterioration of the Practical Virtues. This given, we restricted our attention to probabilistic algorithms.

In Chapters 3 and 4, we argued that mathematicians are under some circumstances rationally obliged to Privately Accept the conclusions of probabilistic algorithms. We gave a definition of a 'Probabilistic Argument' as a discourse suitable for expressing the conclusions yielded by such a method. We then considered relaxing the rule of requiring proof for Public Acceptance, to include results endorsed by any procedure that yields a falsehood with probability less than $\varepsilon = 10^{-10,000}$. We concluded that because of the scope and effectiveness of Monte Carlo techniques, this innovation could be a viable solution to maintaining the Practical Virtues.

However, in Chapter 5 we argued that mathematical research also embodies four corresponding Intellectual Virtues: Abstractness, Explicitness, Univocality and Formalizability. These are shared standards of excellence that partially constrain what a good mathematical discourse can now be. And in Chapter 6, we saw that Probabilistic Arguments cannot be brought into compliance with them.

We have therefore shown that mathematicians can give reasons for rejecting Probabilistic Arguments within the context of Public Acceptance. But are these sufficient reasons? How might we respond to the view that we should sacrifice the

[321] David Hilbert, "Über das Unendliche", *Mathematische Annalen* 95 (1926): 161-190.

Intellectual Virtues in pursuit of the benefits Probabilistic Arguments can bring, both in enhancing the Practical Virtues and for the progress of research in general? For we saw in Section 4.vi that probabilistic methods may provide a viable solution to the challenges to the Practical Virtues raised in Section 1.vi. And there may be kinds of problems that whilst at this time cannot be solved with deductive techniques are amenable to probabilistic methods.

Though mathematicians today are likely to value the Intellectual Virtues for their own sake, there are also two other routes to their rational justification. Firstly, they have direct benefits: we saw in Chapter 5 why adherence to the Intellectual Virtues is important for mathematical enquiry to continue to flourish to the extent it does today. Modern functional analysis would not be possible without a strict adherence to Univocality, for example. Secondly, the Intellectual Virtues were also shown to be important for maintaining the Practical Virtues. For instance, Consensus is difficult to achieve if Univocality and Explicitness are not met, due to the instability of individual intuitive judgement.

The rejoinder may also be countered by consideration of a rival way of protecting the Practical Virtues from threats produced by the nature of contemporary mathematics. In Section 5.vi, we saw that the Probabilistically Checkable Proofs Theorem gives us an immensely powerful way of checking the validity of proofs. Moreover, we noted that the scope of this technique extends across all of Publicly Accepted mathematics: '*any* mathematical theorem, in any standard formal system such as Zermelo-Fraenkel set theory, can be converted in polynomial time into a probabilistically-checkable format.'[322] These checks need not be included as part of Publicly Accepted mathematics proper, and are merely a highly effective way of ascertaining the validity of our deductive arguments.

Making use of this technique requires that our mathematics is Formalizable, which is only possible if the other three Intellectual Virtues are also present. One cannot Formalize an argument containing empirical premisses; if concepts are not unambiguously specified they cannot be identified with suitable formal entities; inferences that essentially rely on complex and irreducible intuitive judgements cannot be rendered as explicit rules in a formal system.

All things considered, then, mathematicians should continue to ensure that their work embodies the Intellectual Virtues. They are therefore not behaving irrationally in continuing to reject Probabilistic Arguments as a basis for Public Acceptance. This is true even though they are under some circumstances required to Privately Accept the results of Monte Carlo algorithms. Such results may even be announced, for example in the *Journal of Experimental Mathematics* – and thus drive the direction of future research. However, these claims must be clearly separated off from those parts of their work they wish to gain Public Acceptance.

[322] Aaronson, "Why Philosophers Should Care About Computational Complexity", 302.

7.ii. Epilogue

Part of our conclusion has been that Public and Private Acceptance may under some circumstances come apart. But this is nothing new and indeed is confirmed by day-to-day mathematical experience. We often rely on a disparate variety of kinds of evidence in coming to believe in our hearts that the answer we have found is correct, and may only then even begin looking for the kind of explicit reasons suitable for Public Acceptance.

Though I still have not checked it deductively, I have argued that I am rationally required to believe the claim that 66,998,713 is prime. The reader is again invited either to accept the claim on the argument given thus far or to generate their own numbers for potential witnesses. In using a probabilistic algorithm known to be highly reliable we have thus arrived, though not by mathematics, at the truth.

Bibliography

Aaronson, Scott. "Why Philosophers Should Care About Computational Complexity". In *Computability: Turing, Gödel, Church, and Beyond*. Edited by B. Jack Copeland, Carl J. Posey, Oron Shagrir. United States: MIT Press, 2013.

Abelson, Hal and Sussman, Gerald J. *Structure and Interpretation of Computer Programs*. Massachusetts: MIT Press, 1996.

Adleman, Leonard. "Molecular Computation of Solutions to Combinatorial Problems". *Science* 266 (1994): 1021-1024.

Almgren, Frederick. *Almgren's Big Regularity Paper*. Edited by Vladimir Scheffer and Jean Taylor. Singapore: World Scientific Publishing, 2000.

Appel, Kenneth and Haken, Wolfgang. "Every Planar Map is Four Colourable. Part I: Discharging". *Illinois Journal of Mathematics* 21 (1977): 429-490.

Appel, Kenneth; Haken,Wolfgang and Koch, John. "Every Planar Map is Four Colourable. Part II: Reducibility". *Illinois Journal of Mathematics* 21 (1977): 491-567.

Aschbacher, Michael and Smith, Stephen. "The Classification of Quasithin Groups: I. Structure of Strongly Quasithin κ-groups". *Mathematical Surveys and Monographs* 111 (2004).

Aschbacher, Michael and Smith, Stephen. "The Classification of Quasithin Groups: II. Main Theorems: The Classification of Simple QTKE-groups". *Mathematical Surveys and Monographs* 112 (2004).

Avigad, Jeremy; Donnelly, Kevin; Gray, David and Raff, Paul; "A Formally Verified Proof of the Prime Number Theorem", *ACM Transactions on Computational Logic* 9 (2007): 1-23

Azzouni, Jodi. "How and Why Mathematics is Unique as a Social Practice". In *18 Unconventional Essays on the Nature of Mathematics*. Edited by Reuben Hersh. New York: Springer, 2006.

Babai, László. "Monte-Carlo Algorithms in Graph Isomorphism Testing". *Technical Report of the DMS* 79 (1979).

Baker, Alan. "Non-deductive methods in mathematics". *Stanford Encyclopedia of Philosophy*. Accessed 10th August 2015. http://plato.stanford.edu/entries/mathematics-nondeductive/

Bauer, Gertrud and Wenzel, Markus. "Computer-Assisted Mathematics at Work (The Hahn-Banach Theorem in Isabelle/Isar)". In *Selected Papers from the*

International Workshop on Types for Proofs and Programs. Edited by Thierry Coquand, Peter Dubjer, Bengy Nordström and Jan M. Smith. London: Springer-Verlag, 2000, 61-76.

Berge, Arno and Hill, Theodore. *An Introduction to Benford's Law*. Princeton: Princeton University Press, 2015.

Berkeley, George. *The Analyst*. In *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*. Edited by William Ewald. Oxford: Oxford University Press.

Berlekamp, Elwyn. "Factoring Polynomials over large finite fields." *Mathematics of Computation* 24 (1970): 713-735.

Bernhart, Frank. "A Digest of the Four Colour Theorem". *Journal of Graph Theory* 1 (1977).

Borwein, Jonathan and Bailey, David. *Mathematics by Experiment: Plausible Reasoning in the 21st Century*. Massachusetts, A K Peters, 2004.

Bottazini, Umberto. *The Higher Calculus*. New York: Springer-Verlag. 1986.

Boyer, Pascal. *History of Analytic Geometry*. New York: Dover, 2004.

Cassirer, Ernest. *The Problem of Knowledge*. Connecticut: Yale University Press, 1978.

Chen Jingrun. "On the Representation of a Large Even Integer as the Sum of a Prime and the Product of at Most Two Primes". *Kexue Tongbao* 11 (1966).

Chudnovsky, Maria; Robertson, Neil; Seymour, Paul and Thomas, Robin. "The Strong Perfect Graph Theorem". Annals of Mathematics 165 (2006): 51-229.

Corfield, David. *Towards a Philosophy of Real Mathematics*. Cambridge: Cambridge University Press, 2004.

Crowe, Michael. *A History of Vector Analysis*. Toronto: University of Notre Dame Press, 1967.

Crupi, Vincenzo. "Confirmation". *Stanford Encyclopedia of Philosophy*. Accessed 17th August 2015. http://plato.stanford.edu/entries/confirmation/

Damgård, Ivan; Landrock, Peter, and Pomerance, Carl. "Average Case Error Estimates for the Strong Probably Prime Test". *Mathematics of Computation* 61 (1993): 177-194.

Davis, Philip. "Fidelity in Mathematical Discourse: Is One and One Really Two?" *The American Mathematical Monthly*, 79 (1972).

Davis, Philip; Hersh, Reuben and Marchisotto, Elena. *The Mathematical Experience: Study Edition*. New York: Berkhäuser, 2012.

De Millo, Richard; Lipton, Richard and Perlis, Alan. "Social Processes and Proofs of Theorems and Programs". *Communications of the ACM* 22 (1979).

De Villiers, Michael. "The Role and Function of Quasi-Empirical Methods in Mathematics". *Canadian Journal of Science, Mathematics and Technology Education* 4, (2004).

Debnath, Lokenath and Bhatta, Dambaru. *Integral Transforms and Their Applications*. London: Chapman and Hall/CRC, 2006.

Deistel, Reinhard. *Graph Theory*. Berlin: Springer-Verlag, 2010.

Dennett, Daniel. *Intuition Pumps and Other Tools for Thinking*. New York: W. W. Norton and Company, 2013.

Descartes, René. *La Geometrié*. In *God Created The Integers*. Edited by Stephen Hawking. London: Penguin, 2004.

Diaconis, Persi; Holmes, Susan and Montgomery, Richard. "Dynamical Bias in the Coin Toss".*SIAM Review* 49 (2007): 211-235.

Dickson, Leonard Eugene. *History of the Theory of Numbers (Two Volumes)*. Chelsea: New York, 1952.

Dretske, Fred. *Knowledge and the Flow of Information*. Cambridge: The MIT Press, 1981.

Easwaran, Kenny. "Probabilistic Proofs and Transferability". *Philosophia Mathematica* 17 (2009): 341-362.

Echeverria, Javier. "Empirical Methods in Mathematics. A Case-Study: Goldbach's Conjecture". In *Spanish Studies in the Philosophy of Science*. Edited by G. Munévar. Boston: Kluwer Academic Publishers, 1996.

Engel, Arthur. *Problem-Solving Strategies*. New York: Springer, 1998.

Euler, Leonhard. "An Inquiry Into Whether or Not 1,000,009 is a Prime Number", *Nova Acta Academiae Scientarum Imperialis Petropolitinae* 10 (1797): 63-7.

Engleman, Steven. *Families of Curves and the Origins of Partial Differentiation*. Amsterdam: Elsevier, 2000.

Euclid. *The Thirteen Books of the Elements, 3 Volumes*. Translated with introduction and commentary by Thomas Heath. New York: Dover, 2012.

Fallis, Don. "The Epistemic Status of Probabilistic Proof". *The Journal of Philosophy* 94 (1997): 165-186.

—— "What do Mathematicians Want? Probabilistic Proofs and the Epistemic Goals of Mathematicians" *Logique et Analyse* 45 (2002): 373-388.

—— "Probabilistic Proofs and the Collective Epistemic Goals of Mathematicians", in *Collective Epistemology*, eds. Hans Bernard Schmid, Marcel Weber, and Daniel Sirtes (Germany: Ontos Verlag, 2011), 157-175.

Fefferman, Charles; Israel, Arie, and Luli Garving. "Sobolev Extension by Linear Operators". *Journal of the American Mathematical Society* 27 (2014).

Fienberg, Stephen. "A Brief History of Statistics in Three and One-half Chapters: A Review Essay". *Statistical Science* 7 (1992): 208–225.

Franklin, James. *The Science of Conjecture: Evidence and Probability before Pascal*. Maryland: The Johns Hopkins University Press, 2001.

Finetti, Bruno de. "Foresight: Its Logical Laws, Its Subjective Sources". *Studies in Subjective Probability*. Edited by H. E. Kyburg, Jr. and H. E. Smokler. New York: Robert E. Krieger Publishing Company, 1980.

—— *Theory of Probability*. New York: Wiley, 1990.

Fowler, David. *The Mathematics of Plato's Academy: A New Reconstruction*. Oxford: The Clarendon Press, 1991.

Frye, Roger. "Finding $95800^4 + 217519^4 + 414560^4 = 422481^4$ on the Connection Machine". *Proceedings of Supercomputing* 88 (1988): 106-116.

Frege, Gottlob. *Foundations of Arithmetic*. Translated by John Austin. Oxford: Basil Blackwell, 1980.

Gettier, Edmund. "Is Justified True Belief Knowledge?" *Analysis* 23 (1963).

Giaquinto, Marcus. *The Search for Certainty: A Philosophical Account of Foundations of Mathematics*. Oxford: Oxford University Press, 2002.

Gödel, Kurt. "Über die Vollständigkeit des Logikkalküls". Doctoral Dissertation. University of Vienna, 1929.

Goldberg, Michael. "On the Original Malfatti Problem". *Mathematics Magazine*, 40 (1967): 241-247.

Golin, Mordecai; Raman, Rajeev; Schwarz, Christian, and Smid, Michiel. "Randomized Data Structure for the Dynamic Closest Pair Problem". *Society for Industrial and Applied Mathematics*, 27 (1998): 1036-1072.

Golomb, Solomon. "A Combinatorial Proof of Fermat's Little Theorem". *The American Mathematical Monthly* 63 (1956).

Gonthier, Georges. "Formal Proof – The Four-Color Theorem". *Notices of the AMS* 55 (2008): 1382-1393.

Gorensten, Daniel; Lyons, Richard and Solomon, Ronald. "The Classification of the Finite Simple Groups, Part 1". *American Mathematical Society Surveys and Monographs* 40.1

Gorroochurn, Prakash. "Some Laws and Problems of Classical Probability and How Cardano Anticipated Them". *Chance* 25 (2012): 13–20.

Grabiner, Judith. "Is Mathematical Truth Time-Dependent?" *The American Mathematical Monthly* 81 (1974).

Grabiner, Judith. "Who gave you the epsilon? Cauchy and the origins of rigorous calculus". *The American Mathematical Monthly*, 90 (1983): 185– 194.

Grünbaum, Branko. "Quadrangles, Pentagons, and Computers". *Geombinatorics* 3 (1993).

Hájek, Alan. "Fifteen Arguments Against Hypothetical Frequentism". *Erkenntnis* 70 (2009): 211-235

—— "Intepretations of Probability". *Stanford Encyclopedia of Philosophy*. Accessed 22nd August 2015. http://plato.stanford.edu/entries/probability-interpret/

Hales, Thomas. "A Proof of the Kepler Conjecture". *Annals of Mathematics* 162 (2005): 1065-1185.

Halmos, Paul. "Mathematics as a creative art". In *Mathematics: People, Problems, Results, Vol 2*. California: Wadsworth, 1984.

Harada, Koichiro and Solomon, Ronald. "Finite groups having a standard component L of type $\widehat{M}_{12}$ or $\widehat{M}_{22}$". *Journal of Algebra* 319 (2008), 621–628.

Hardy, Godfrey H. *Ramanujan*. New York: Cambridge University Press, 1940.

Hardy, Godfrey H. *A Mathematician's Apology*. Cambridge: Cambridge University Press, 1992.

Hardy, Godfrey H. and Littlewood, John. "On Some Problems of "Partitio Numerorum" III: On the Expression of a Number as the Sum of Primes", *Acta Mathematica* 44 (1923): 1-70.

Haselgrove, Colin. "A Disproof of a Conjecture of Pólya." *Mathematika* 5 (1958), 141-145.

Hawthorne, John. *Knowledge and Lotteries*. New York: Oxford University Press, 2004.

Headwood, Percy. "Map-Colour Theorem". *Quarterly Journal of Mathematics,* 24 (1890): 332–338.

Helfgott, Harald. “The Ternary Goldbach Conjecture is True”. 2013. arXiv:1312.7748

Hilbert, David. “Über das Unendliche.” *Mathematische Annalen* 95 (1926): 161-190.

Hoare, Sir Charles ‘Tony’. “Algorithm 64: Quicksort”. *Communications of the ACM* 4 (1961): 321.

Hume, David. *A Treatise of Human Nature*. Oxford: Oxford University Press, 2000.

Jaffe, Arthur Jaffe and Quinn, Frank. ““Theoretical Mathematics”: Toward A Cultural Synthesis Of Mathematics And Theoretical Physics”. *Bulletin of the American Mathematical Society* 29 (1993).

Jourdain, Philip. “The Nature of Mathematics”. In *The World of Mathematics*, *Vol 1*. Edited by James Newman. New York: Dover, 1956.

Karp, Richard. “An Introduction to Randomized Algorithms”. *Discrete Applied Mathematics* 34 (1991): 165-201.

Kahneman, Daniel. *Thinking Fast and Slow*. London: Penguin, 2011.

Kahneman, Daniel and Tversky, Amos. “Judgement under Uncertainty: Heuristics and Biases”. *Science* 185 (1974): 1124-1131.

Katz, Victor J. “Euler’s Analysis Textbooks”. In *Leonard Euler: Life, Work and Legacy*. Edited by Robert Bradley and C. Edward Sandifer. Oxford: Elsevier, 2007.

Keller, Joseph B. “The Probability of Heads”. *The American Mathematical Monthly* 93 (1986): 191-197.

Knorr, Wilbur Richard. *The Ancient Tradition of Geometric Problems*. New York: Dover, 1993.

Kolmogorov, Andrey. *Foundations of the Theory of Probability*. New York: Chelsea, 1956.

Korniłowicz, Artur. “A Proof of the Jordan Curve Theorem via the Brouwer Fixed Point Theorem”. *Mechanized Mathematics and its Application* 6 (2007): 33-40.

Knuth, Donald. *The Art of Computer Programming*. Addison-Wesley: Massachusetts, 1981.

Kuhn, Thomas. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1996.

Imre Lakatos, *Proofs and Refutations* (Cambridge: Cambridge University Press, 1976)

Lamé, Gabriel. "*Démonstration generale du théorème de Fermat*". *Compte Rendu des Séances de L'Academie des Science* (1847): 310-315.

Lander, L. J. and Parkin, T. R. "Counterexample to Euler's Conjecture on Sums of Like Powers". *Bulletin of the American Mathematical Society* 72 (1966).

Laplace, Pierre Simon de. *Concerning Probability*. In *The World of Mathematics, Vol 1*. Edited by James Newman. New York: Dover, 1956.

Laumon, Gérard. "The Work of Laurent Lafforgue". *International Congress of Mathematicians* 1 (2002).

Lehman, Russell. "On Liouville's Function". *Mathematics of Computation* 14 (1960): 311-320.

Lewis, David. "Humean Supervenience Debugged". *Mind* 103 (1994): 473-490

Lob, H. and Richmond, H. W. "On the Solutions of Malfatti's Problem for a Triangle". *Proceedings of the London Mathematical Society* 30 (1930): 287-304.

Lützen, Jesper. "Between Rigor and Applications: Developments in the Concept of Function in Mathematical Analysis". In *Cambridge History of Science, Volume 5*. Edited by Mary Jo Nye.

Maanen, Jan van. "Precursors of Differentiation and Integration". In *A History of Analysis*. Idited by Hans Niels Jahnke. London: London Mathematical Society, 2003.

Macintyre, Alasdair. *After Virtue*. London: Bloomsbury, 2011.

—— "On Having Survived the Academic Moral Philosophy of the Twentieth Century". In *What Happened in and to Moral Philosophy in the Twentieth Century*. Edited by Fran O'Rourke. Notre Dame: Notre Dame University Press, 2013.

Malfatti, Gian Francesco. "Memoria sopra un problema sterotomico". *Memorie di Maternatica e di Fisica della Societa Italiana delle Scienze* 10 (1803): 235-244.

Mancosu, Paolo. "Visualization in Logic and Mathematics". In *Visualization, Explanation and Reasoning Styles*. Edited by Paolo Mancosu, Klaus Frovin Jørgensen, and Stig Andur Pedersen. Dordrecht: Springer, 2005.

Manindra, Agrawal; Kayal, Neeraj and Saxena, Nitin. "PRIMES is in P". *Annals of Mathematics* 160 (2004): 781-793.

McCune, William. “Solution of the Robbins Problem”. *Journal for Automated Reasoning* 19 (1997): 263-276.

Motwani, Rajeev and Raghaven, Prabhakar. Randomized Algorithms. Cambridge: Cambridge University Press, 1995.

Miller, Gary. “Riemann’s Hypothesis and Tests for Primality”. *Journal of Computer and System Sciences* 13 (1976): 300-317.

Mitchem, John. “On the History and Solution of the Four-Color Map Problem”. *The College Mathematics Journal* 12 (1981), 108-116.

Mullen, Gary L. and Panario, Daniel. *Handbook of Finite Fields*. Florida: CRC Press, 2003.

Narkiewicz, Władysław. *The Development of Prime Number Theory: From Euclid to Hardy and Littlewood*. New York: Springer-Verlag, 2000.

Neeman, Amnon. “A Counterexample to a 1961 “Theorem” in Homological Algebra”. *Inventiones Mathematicae* 148 (2002): 397-420.

Newton, Isaac. “The October 1666 Tract on Fluxions”. In *The Mathematical Papers of Isaac Newton*, edited by D. T. Whiteside (Cambridge: Cambridge University Press, 1967).

Odlyzko, Andrew and Riele, Herman te. “Disproof of the Mertens Conjecture”. *Journal für die reine und angewandte Mathematik* 357 (1985): 138-160.

Ramsey, Frank. “Truth and Probability”. *Foundations of Mathematics and Other Logical Essays*. Abingdon: Routledge, 2006.

Sabbagh, Karl. *The Riemann Hypothesis*. New York: Farrar, Strauss and Giroux, 2003.

Silva, Tomás Oliveira e; Herzog, Siegfried and Pardi, Silvio. “Empirical Verification of the Even Goldbach Conjecture and Computation of Prime Gaps up to 4×10^{18}”. *Mathematics of Computation* 83 (2013): 2033-2060.

Singh, Simon. *Fermat’s Last Theorem*. London: Harper Perennial, 2002.

Paulson, Larry “A Machine-Assisted Proof of Gödel’s Incompleteness Theorems for the Theory of Hereditary Finite Sets”. *Review of Symbolic Logic* 7 (2014), 484-498.

Peirce, Charles Sanders. “Note on the Doctrine of Chances”. In *Collected Papers of Charles S. Peirce*. Charles Harteshorne, Paul Weiss, and Arthur Burks. Massachusetts: Harvard University Press, 1931-1958.

—— “The Red and the Black”. In *The World of Mathematics*. Edited by James Newman. New York: Dover, 1956.

Pierpont, James. “On the Arithmetization of Mathematics”. *Bullitin of the American Mathematical Society* 5 (1899).

Plato. *The Republic*. Translated by Tom Griffith. Edited by G. R. F. Ferrari. Cambridge: Cambridge University Press, 2007.

Pledger, Keith and Wilkins, Dave. *Edexcel AS and A Level Modular Mathematics, C2*. Portsmouth: Heinemann, 2008.

Pólya, George. “Verschiedene Bemerkungen zur Zahlentheorie”. *Jahresber Deutschen Math.-Verein* 28 (1919): 31-40.

—— *Mathematics and Plausible Reasoning, Volume I*. Princeton: Princeton University Press, 1954.

Pomerance, Carl and Crandal, Richard. *Prime Numbers: A Computational Perspective*. New York: Springer, 2001.

Popper, Karl. “The Propensity Interpretation of Probability”. *The British Journal for the Philosophy of Science* 10 (1959), 25-42.

Proclus. *A Commentary on the First Book of Euclid’s Elements*. Translated with an introduction by Glenn Morrow. Princeton: Princeton University Press, 1992.

Rabin, Michael. “Probabilistic Algorithms”. in *Algorithms and Complexity: New Directions and Recent Trends*. Edited by J. F Traub. New York: Academic Press, 1976. 21-39.

Rabin, Michael. “Probabilistic Algorithm for Primality Testing”. *Journal of Number Theory* 12 (1980): 128-138.

Reich, Karin. “Die Entdeckung und frühe Rezeption der Konstruierbarkeit des regelmäβigen 17-Ecks und dessen geometrische Konstruktion durch Johannes Erchinger (1825)”. In *Mathesis. Festschrift zum siebzigsten Geburtstag von Matthias Schramm*. Edited by R. Thiele Berlin: Diepholz, 2000. 101-118

Reid, Constance. *Hilbert*. New York: Springer, 1996.

Ribet, Ken. “On Modular Representations of $Gal(\bar{Q}/Q)$ Arising From Modular Forms”. *Inventiones Mathematicae* 100 (1990): 431-476

Rokicki, Tomas; Kociemba, Herbert; Davidson, Morley and Dethridge, John. “The Diameter of the Rubik’s Cube Group is Twenty”. *Siam Journal on Discrete Mathematics* 27 (2013): 1082-1105.

Ross, Fiona and Ross, William. “The Jordan Curve Theorem is Non-Trivial.” *Journal of Mathematics and the Arts* 0 (2009): 1-4.

Rosser, Barkley. “Explicit Bounds for some functions of prime numbers”. *American Journal of Mathematics*, 63 (1941).

Rotman, Joseph. *Journey into Mathematics: An Introduction to Proofs*. New Jersey: Prentice Hall, 1998.

Russell, Bertrand. “A Liberal Decalogue”. In *Autobiography*. London: Routledge, 2009.

Savage, Leonard. *The Foundations of Statistics*. New York: Dover, 1972.

Schwartz, Jack. “Fast Probabilistic Algorithms for Verification of Polynomial Identities”. *Journal of the ACM* 27 (1980): 701-717.

Seress, Ákos. *Permutation Group Algorithms*. Cambridge: Cambridge University Press, 2003.

Solomon, Roland. “A Brief History of the Classification of Finite Simple Groups”. *Bulletin of the American Mathematical Society* 22 (2001)

Solomon, Roland. “On Finite Simple Groups and Their Classification”. *Notices of the AMS* 42 (1995).

Solovay, Robert and Strassen, Volker. “A Fast Monte-Carlo Test for Primality”. *SIAM Journal on Computing* 6 (1977): 84-85.

Stein, Elias and Shakarachi, Rami. *Complex Analysis*. Princeton: Princeton University Press, 2003.

Talbott, William. “Bayesian Epistemology”. *Stanford Encyclopedia of Philosophy*. Accessed 11th August 2015. http://plato.stanford.edu/entries/epistemology-bayesian/

Tall, David. “The Nature of Mathematical Proof”. *Mathematics Teaching* 127 (1989)

Tanaka, Minoru. “A Numerical Investigation on Cumulative Sum of the Liouville Function”. *Tokyo Journal of Mathematics* 3 (1980): 187-189.

Thurston, William. “On Proof And Progress In Mathematics”. *For the Learning of Mathematics* 15 (1995).

Tymoczko, Thomas. “The Four-Colour Problem and its Philosophical Significance”. *The Journal of Philosophy* 76 (1979): 57-83.

Ulam, Stanislaw. *Adventures of a Mathematician*. Oakland: University of California Press, 1992.

Venn, John. *The Logic of Chance*. New York: Chelsea Publishing Co., 1962.

Weierstrass, Karl. "On Continuous Functions of a Real Argument that do not Possess a Well-Defined Derivative for any Value of their Argument." In G. A. Edgar, *Classics on Fractals*. Boston: Addison-Wesley Publishing Company: 1993.

Whitehead, Alfred North. "Mathematics as an Element in the History of Thought." In *The World of Mathematics*, *Vol 1*. Edited by James Newman. New York: Dover, 1956.

Wigner, Eugene. "The Unreasonable Effectiveness of Mathematics in the Natural Sciences." *Communications on Pure and Applied Mathematics* 13 (1960): 1-14.

Wiles, Andrew. "Modular Elliptic Curves and Fermat's Last Theorem". *Annals of Mathematics* 142 (1995): 443-551.

Whewell, William. *Lectures on the History of Moral Philosophy*. Bristol: Thoemmes, 1990.

Wittgenstein, Ludwig. *Lectures on the Foundations of Mathematics*. Chicago: University of Chicago Press, 1989.

Yohe, J. M. "Computer Programming for Accuracy". *Proceedings of the 1968 Army Numerical Analysis Conference, U. S. Army Research Office, Durham, North Carolina* (1968).

Youschkevich, A. P. "The Concept of Function up to the Middle of the 19th Century." *Archive for History of Exact Sciences* 16 (1976): 37-85.

Zeilberger, Doron. "Theorems for a Price". *Notices of the AMS* 40 (1993): 978-981.

Zeilberger, Doron and Wilf, Herb. "An Algorithmic Proof Theory for Hypergeometric (Ordinary and 'q') Multisum/Integral Identities". *Inventiones Mathematicae* 108 (1992): 575-633.

Zermelo, Ernst. "Untersuchungen über die Grudnlagen de Mengenlehre". *Mathematische Annalen* I (1908): 261-281

Ziegler, James and Langford, William. "Effect of Cosmic Rays on Computer Memories". *Science* 16 (1979), 776-788.

Zippel, Richard. "Probabilistic Algorithms for Sparse Polynomials". In *Symbolic and Algebraic Computation*. Edited by Edward Ng. New York: Springer, 1979: 216-226